

	10	28.8	36.0	42.14	22	AAH17831	Human CDNA sequen
	11	28.8	36.0	67.16	22	ABA20474	Human nervous syst
	12	28.8	36.0	77.20	22	ABA20475	Human nervous syst
	13	28.4	35.5	108.07	16	AATU03855	Sheep beta-lactogl
	14	28.4	35.5	108.07	18	AATV79725	Ovine beta-lactogl
c	15	28.2	35.2	1636	23	ABL037895	Drosophila melanog
c	16	28.2	35.2	2028	23	ABL037865	Drosophila melanog
c	17	28.2	35.2	7564	23	ABL03788	Drosophila melanog
	18	28.2	35.2	8482	23	ABL19864	Drosophila melanog
	19	27.6	34.5	222	18	AAI60633	Drosophila melanog
	20	27.6	34.5	824	21	AAZ87633	Human beta-lactogl
	21	27.6	34.5	824	21	AAC68328	Sheep beta-lactogl
c	22	27.4	34.2	879	21	AAF16035	Human prostate can
c	23	27.4	34.2	31813	22	ABA08137	Human ovarian and
	24	27.4	34.2	31813	22	AAI06997	Human reproductive
	25	26.8	33.5	4094	21	AAE531205	Human hypoxia regu
	26	26.8	33.5	4248	21	AAA28450	Human Scladin-1 CD
c	27	26.6	33.2	2712	21	AAIC46769	Arabidopsis thaliaa
c	28	26.4	33.0	2780	23	ABLI13960	Drosophila melanog
c	29	26.4	33.0	2809	22	ABLI19898	Drosophila melanog
c	30	26.4	33.0	3552	22	AAK51875	Human polynucleoti
c	31	26.4	33.0	3679	22	AAK52859	Human polynucleoti
c	32	26.4	33.0	7461	22	AAAS30639	Human encoding novel
c	33	26.4	33.0	7461	22	AAAS28701	Genomic sequence #
c	34	26.4	33.0	32174	22	ABAI15665	Human nervous syst
c	35	26.4	33.0	32174	22	ABAI19477	Human nervous syst
c	36	26.4	33.0	32174	22	ABA203305	Human nervous syst
c	37	26.4	33.0	32174	22	ABA215055	Human nervous syst
c	38	26.4	33.0	32174	22	AAI36280	Human musculoskele
c	39	26.4	33.0	32174	22	AAAS36555	Human genomic DNA
c	40	26.4	33.0	32174	22	AAAS36394	Human DNA for a no
c	41	26.4	33.0	32174	22	AAAS30638	DNA encoding novel
c	42	26.4	33.0	32174	22	AAAL03792	Human reproductive
c	43	26.4	33.0	32174	22	AAI074447	Human reproductive
c	44	26.4	33.0	32174	22	AAI074491	Human reproductive
c	45	26.4	33.0	32174	22	AAAS28700	Genomic sequence #

ALIGNMENTS

```

RESULT_1
ID      AAA28495 standard; DNA; 80 BP.
XX AC   AAA28495;
XX DT   29-AUG-2000 (first entry)
DE O. aureus vitellogenin secretory signal sequence coding sequence.
XX KW   Vitgss; vitellogenin; secretory signal sequence; gene expression;
XX KM   oestrogen receptor binding protein; systemic circulation; ss.
XX OS   Oreochromis aureus.
FH Key Location/Qualifiers
FT sig_peptide 18..80
FT FT          /*tag= a
PN WO200026366-A1.
PD
XX PD
XX PF   29-OCT-1999; 99WO-SG00108.
XX PR   30-OCT-1998; 98US-0106426.
XX PR   26-OCT-1999; 99US-0426776.
XX PA   (UYST-) UNIV SINGAPORE NAT.
XX PA   (LAMT/) LAM T J.
XX Ding JL, Tan NS, Ho B;
```

```

XX  WPI; 2000-365615/31.
DR  P-PSDB; AAY92778.
XX
XX  Isolated nucleic acid for assaying for heterologous gene expression,
PT  detecting presence of compound that binds to estrogen receptor or
PT  producing desired protein from host cell comprises nucleotide sequence
PT  encoding secretory signal sequence
XX
XX  Claim 2; Page 36; 73pp; English.
XX
XX  The present sequence encodes a piscine, Oreochromis aureus, vitellogenin
CC  secretory sequence (Vtgs). This and variants that comprise conservative
CC  replacements that retain the biological activities of directing secretion
CC  of a fusion protein from a cell and cleavage of the secretory signal
CC  sequence from the fusion protein, are new. DNA encoding the Vtgs can be
CC  fused to either a reporter protein or a lipopolysaccharide-binding
CC  protein coding sequence. The isolated nucleic acid is useful for assaying
CC  for heterologous gene expression, detecting the presence of a compound
CC  that binds to an estrogen receptor in a sample or producing a desired
CC  protein from a host cell. It can also be used in a method for obtaining
CC  systemic circulation of a desired protein in a transgenic or chimeric
CC  host organism.
XX
SQ  Sequence 80 BP; 14 A; 22 C; 24 G; 20 T; 0 other;

Query Match      100.0%; Score 80; DB 21; Length 80;
Best Local Similarity 100.0%; Pred. No. 1.7e-17;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 attcacatccaccagcagatgaggtgtgtactagcttctgtgtgtctgcagatg 60
    |||||||
DB  1 attcacatccaccagcagatgaggtgtgtactagcttctgtgtgtctgcagatg 60
    |||||||
OY  61 ggaccagtcacacttgagg 80
    |||||||
DB  61 ggaccagtcacacttgagg 80
    |||||||

RESULT  2
ID  AAA28499 standard; DNA; 152 BP.
XX
XX  AAA28499;
AC  AAA28499;
XX
XX  29-AUG-2000 (first entry)
DE  Vtgs-CAT fusion protein (partial) coding sequence.
XX
XX  Vtgs; vitellogenin; secretory signal sequence; gene expression;
KM  oestrogen receptor binding protein; systemic circulation; CAT; ss.
XX
XX  Chimeric - Oreochromis aureus.
OS  Chimeric - Synthetic.
XX
XX  Location/Qualifiers
FH  Key
FT  CDS
FT  51..152
FT  /tag= a
FT  /transl_except= (pos:123..125, aa:Gln)
FT  /transl_except= (pos:132..134, aa:Val)
FT  /transl_except= (pos:141..143, aa:Trp)
FT  sig_peptide
FT  51..113
FT  /*tag= b
FT  117..152
FT  /*tag= c
FT  /partial
FT  /note= "CAT N-terminal"
XX
XX  WO200026366-A1.
XX
XX  11-MAY-2000.

```

```

PF  29-OCT-1999; 99MO-SG00108.
XX
XX  30-OCT-1998; 98US-0106426.
PR  26-OCT-1999; 99US-0426776.
XX
XX  (UYSI-) UNIV SINGAPORE NAT.
PA  (LAWT/) LAW T J.
XX
XX  Ding JL, Tan NS, Ho B;
PI  WPI; 2000-365615/31.
XX  P-PSDB; AAY92780.
XX
XX  Isolated nucleic acid for assaying for heterologous gene expression,
PT  detecting presence of compound that binds to estrogen receptor or
PT  producing desired protein from host cell comprises nucleotide sequence
PT  encoding secretory signal sequence
XX
XX  Example 2; Fig 5B; 73pp; English.
XX
XX  A reporter CAT system that uses the piscine, Oreochromis aureus,
CC  vitellogenin secretory sequence (Vtgs), designated pBSVtgsCAT was
CC  constructed. AAA28499 and AAA28722 comprise the 5'- and 3'-ends of the
CC  construct insert. Vtgs and variants that comprise conservative
CC  replacements that retain the biological activities of directing
CC  secretion of a fusion protein from a cell and cleavage of the secretory
CC  signal sequence from the fusion protein, are new. DNA encoding the Vtgs
CC  can be fused to either a reporter protein or a
CC  lipopolysaccharide-binding protein coding sequence. The isolated nucleic
CC  acid is useful for assaying for heterologous gene expression, detecting
CC  the presence of a compound that binds to an estrogen receptor in a
CC  sample or producing a desired protein from a host cell. It can also be
CC  used in a method for obtaining systemic circulation of a desired protein
CC  in a transgenic or chimeric host organism.
XX
SQ  Sequence 152 BP; 38 A; 38 C; 39 G; 37 T; 0 other;

Query Match      100.0%; Score 80; DB 21; Length 152;
Best Local Similarity 100.0%; Pred. No. 1.9e-17;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY  1 attcacatccaccagcagatgaggtgtgtactagcttctgtgtgtctgcagatg 60
    |||||||
DB  34 attcacatccaccagcagatgaggtgtgtactagcttctgtgtgtctgcagatg 93
    |||||||
OY  61 ggaccagtcacacttgagg 80
    |||||||
DB  94 ggaccagtcacacttgagg 113
    |||||||

RESULT  3
ID  AAA28500 standard; DNA; 152 BP.
XX
XX  AAA28500;
AC  AAA28500;
XX
XX  29-AUG-2000 (first entry)
DE  Vtgs-CAT fusion protein (partial) coding sequence.
XX
XX  Vtgs; vitellogenin; secretory signal sequence; gene expression;
KM  oestrogen receptor binding protein; systemic circulation; CAT; ss.
XX
XX  Chimeric - Oreochromis aureus.
OS  Chimeric - Synthetic.
XX
XX  Location/Qualifiers
FH  Key
FT  CDS
FT  51..152
FT  /tag= a
FT  /transl_except= (pos:123..125, aa:Gln)
FT  /transl_except= (pos:132..134, aa:Val)
FT  /transl_except= (pos:141..143, aa:Trp)

```

```

FT sig_peptide 51..113
FT mat_peptide /*tag- b
FT 117..152 /*tag- c
FT /partial
FT /note= "CAT N-terminal"
PW0200026366-A1.
11-MAY-2000.
29-OCT-1999; 99WO-SG00108.
30-OCT-1998; 98US-0106426.
26-OCT-1999; 99US-0426776.
XX (UYSI-) UNIV SINGAPORE NAT.
XX (LAMT/) LAM T J.
XX Ding JL, Tan NS, Ho B;
XX MPI: 2000-365615/31.
XX P-PSDB: AAY92780.
XX Isolated nucleic acid for assaying for heterologous gene expression,
XX detecting presence of compound that binds to estrogen receptor or
XX producing desired protein from host cell comprises nucleotide sequence
XX encoding secretory signal sequence
XX
XX Example 2; Fig 5D; 73pp; English.
XX
XX A reporter CAT system that uses the pISCNE, Oreochromis aureus,
XX vitellogenin secretory sequence (Vtgs), designated pSP-VtgsCAT was
XX constructed. AAA28500 and AAA28723 comprise the 5'- and 3'-ends of the
XX construct insert. Vtgs and variants that comprise conservative
XX replacements that retain the biological activities of directing
XX secretion of a fusion protein from a cell and cleavage of the secretory
XX signal sequence from the fusion protein, are new. DNA encoding the Vtgs
XX can be fused to either a reporter protein or a
XX lipopolysaccharide-binding protein coding sequence. The isolated nucleic
XX acid is useful for assaying for heterologous gene expression, detecting
XX the presence of a compound that binds to an estrogen receptor in a
XX sample or producing a desired protein from a host cell. It can also be
XX used in a method for obtaining systemic circulation of a desired protein
XX in a transgenic or chimeric host organism.
XX
XX Sequence 152 BP; 39 A; 38 C; 39 G; 36 T; 0 other;
SQ
Query Match 100.0%; Score 80; DB 21; Length 152;
Best Local Similarity 100.0%; Pred. No. 1.9e-17;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 attcaatccaccagcagtgagtgctgtactagctcttctgtgctcgcagctg 60
DB 34 attcaatccaccagcagtgagtgctgtactagctcttctgtgctcgcagctg 93
OY 61 gggaccagtcacaacttg99g 80
DB 94 gggaccagtcacaacttg99g 113
RESULT 4
AAA28502
ID AAA28502 standard; DNA; 155 BP.
XX
XX AAA28502;
XX
XX 29-AUG-2000 (first entry)
XX
XX Vtgs-BGFP fusion protein (partial) coding sequence.
XX
XX Vtgs; vitellogenin; secretory signal sequence; gene expression;
KW

```

```

KW oestrogen receptor binding protein; systemic circulation; EGFP; ss.
XX
XX Chimeric - Oreochromis aureus.
XX
XX Chimeric - Synthetic.
XX
XX Key Location/Qualifiers
XX CDS 42..155
XX sig_peptide /*tag- a
XX 42..104
XX /*tag- b
XX /note= "O. aureus Vtgs"
XX mat_peptide /*tag- c
XX 123..155
XX /note= "EGFP"
PW0200026366-A1.
11-MAY-2000.
29-OCT-1999; 99WO-SG00108.
30-OCT-1998; 98US-0106426.
26-OCT-1999; 99US-0426776.
XX (UYSI-) UNIV SINGAPORE NAT.
XX (LAMT/) LAM T J.
XX Ding JL, Tan NS, Ho B;
XX MPI: 2000-365615/31.
XX P-PSDB: AAY92782.
XX Isolated nucleic acid for assaying for heterologous gene expression,
XX detecting presence of compound that binds to estrogen receptor or
XX producing desired protein from host cell comprises nucleotide sequence
XX encoding secretory signal sequence
XX
XX Example 3; Fig 8B; 73pp; English.
XX
XX A reporter GFP system that uses the pISCNE, Oreochromis aureus,
XX vitellogenin secretory sequence (Vtgs), designated pVtgsGFP was
XX constructed. Vtgs and variants that comprise conservative
XX replacements that retain the biological activities of directing secretion
XX of a fusion protein from a cell and cleavage of the secretory signal
XX sequence from the fusion protein, are new. DNA encoding the Vtgs can be
XX fused to either a reporter protein or a lipopolysaccharide-binding
XX protein coding sequence. The isolated nucleic acid is useful for assaying
XX for heterologous gene expression, detecting the presence of a compound
XX that binds to an estrogen receptor in a sample or producing a desired
XX protein from a host cell. It can also be used in a method for obtaining
XX systemic circulation of a desired protein in a transgenic or chimeric
XX host organism.
XX
XX Sequence 155 BP; 30 A; 45 C; 49 G; 31 T; 0 other;
SQ
Query Match 100.0%; Score 80; DB 21; Length 155;
Best Local Similarity 100.0%; Pred. No. 1.9e-17;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 attcaatccaccagcagtgagtgctgtactagctcttctgtgctcgcagctg 60
DB 25 attcaatccaccagcagtgagtgctgtactagctcttctgtgctcgcagctg 84
OY 61 gggaccagtcacaacttg99g 80
DB 85 gggaccagtcacaacttg99g 104
RESULT 5
AAA28496
ID AAA28496 standard; DNA; 204 BP.
XX

```

AC AAA28496;
 XX
 DT 29-AUG-2000 (first entry)
 XX
 DE Vtgs-CRCEs fusion protein coding sequence.
 XX
 KW Vtgs; vitellogenin; secretory signal sequence; gene expression;
 XX oestrogen receptor binding protein; systemic circulation; factor C; ss.
 XX
 OS Chimeric - Oreochromis aureus.
 OS Chimeric - Carinioscorpius rotundicauda.
 FH
 FH Key Location/Qualifiers
 CDS 52..204
 FT /*tag- a
 FT /transl_except- (pos:199..201, aa:Tyr)
 FT /transl_except- (pos:202..204, aa:Phe)
 FT /partial
 FT 52..114
 FT /*tag- b
 FT /note- "O. aureus vitellogenin secretory signal sequence"
 FT 115..204
 FT /*tag- c
 FT /note- "C. rotundicauda Factor C LPS-binding domain"
 XX
 PN W0200026366-A1.
 XX
 PD 11-MAY-2000.
 XX
 PF 29-OCT-1999; 99WO-SG00108.
 XX
 PR 30-OCT-1998; 98US-0106426.
 PR 26-OCT-1999; 99US-0426776.
 XX
 PA (UYSI-) UNIV SINGAPORE NAT.
 PA (LAMT/) LAM T J.
 PI Ding JL, Tan NS, Ho B;
 PI WPI: 2000-365615/31.
 DR P-PSDB: AAY92779.
 XX
 XX Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence
 XX
 PS Example 1; Fig 2A; 73pp; English.
 XX
 CC The piscine, Oreochromis aureus, vitellogenin secretory sequence (Vtgs)
 CC was fused upstream to the EcoRI-SalI CDNA fragment encoding the
 CC lipopolysaccharide-binding domain of Carinioscorpius rotundicauda Factor
 CC C CRCEs for expression and secretion of recombinant ES protein from
 CC Drosophila cells. Vtgs and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an oestrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.
 XX
 XX Sequence 204 BP; 48 A; 45 C; 58 G; 53 T; 0 other;

DB 35 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 35 attacatccaccagcatgaggtgcttctgactagctcttgcgtgctctgcagtg 94
 QY 61 gggaccagtcacttggg 80
 ||||||||||||||||||||||||
 DB 95 gggaccagtcacttggg 114
 ||||||||||||||||||||||||
 RESULT 6
 AAA28507
 ID AAA28507 standard; DNA: 204 BP.
 XX
 XX AAA28507;
 AC
 XX
 DT 29-AUG-2000 (first entry)
 XX
 DE Vtgs-beta-lactamase fusion protein (partial) coding sequence.
 XX
 KW Vtgs; vitellogenin; secretory signal sequence; gene expression;
 KW oestrogen receptor binding protein; systemic circulation;
 KW beta-lactamase; ss.
 XX
 OS Chimeric - Oreochromis aureus.
 OS Chimeric - Synthetic.
 FH
 FH Key Location/Qualifiers
 CDS 52..204
 FT /*tag- a
 FT /partial
 FT 52..114
 FT /*tag- b
 FT /note- "Vtgs from O. aureus"
 FT 115..204
 FT /*tag- c
 FT /note- "beta-lactamase mature protein"
 XX
 PN W0200026366-A1.
 XX
 PD 11-MAY-2000.
 XX
 PF 29-OCT-1999; 99WO-SG00108.
 XX
 PR 30-OCT-1998; 98US-0106426.
 PR 26-OCT-1999; 99US-0426776.
 XX
 PA (UYSI-) UNIV SINGAPORE NAT.
 PA (LAMT/) LAM T J.
 PI Ding JL, Tan NS, Ho B;
 PI WPI: 2000-365615/31.
 DR P-PSDB: AAY92783.
 XX
 XX Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence
 XX
 PS Example 6; Fig 14A; 73pp; English.
 XX
 CC A reporter beta-lactamase system that uses the Oreochromis aureus,
 CC vitellogenin secretory sequence (Vtgs), designated pBAVtgb-lactKana was
 CC constructed. Vtgs and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an oestrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.

Query Match 100.0%; Score 80; DB 21; Length 204;
 Best Local Similarity 100.0%; Prod. No. 2e-17;
 Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 attacatccaccagcatgaggtgcttctgactagctcttgcgtgctctgcagtg 60

XX	SO	Sequence	204 BP; 49 A; 42 C; 60 G; 53 T; 0 other;
XX	Query Match	81.8%; Score 65.4; DB 21; Length 204;	
XX	Best Local Similarity	98.5%; Pred. No. 1.5e-12;	
XX	Matches	66; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
QY	14	accacatgaggtctgtctactagctctctgtgtgtgctctgcagctggggaccagctccaa 73	
Db	48	aacctgaggggtgtctgtactagctctctgtgtgtgctctgcagctggggaccagctccaa 107	
OY	74	cttgagg 80	
Db	108	cttgagg 114	
RESULT 7			
AAH11590/c			
ID	AAH11590	standard; cDNA; 474 BP.	
XX	AC	AAH11590;	
XX	DT	26-JUN-2001 (first entry)	
XX	DE	Human cDNA clone (3'-primer) SEQ ID NO:8425.	
XX	KM	Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.	
XX	OS	Homo sapiens.	
XX	PN	EP1074617-A2.	
XX	PD	07-FEB-2001.	
XX	PF	28-JUL-2000; 2000EP-0116126.	
XX	PR	29-JUL-1999; 99JP-0248036.	
XX	PR	27-AUG-1999; 99JP-0300253.	
XX	PR	11-JAN-2000; 2000JP-0118776.	
XX	PR	02-MAY-2000; 2000JP-0183767.	
XX	PR	09-JUN-2000; 2000JP-0241899.	
XX	PA	(HELI-) HELIX RES INST.	
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;		
PI	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;		
DR	WPI; 2001-318749/34.		
PS	Claim 3; SEQ ID 8425; 2537bp + CD ROM; English.		
XX	XX		
CC	XX	The present invention describes primer sets for synthesizing 5602	
CC	XX	full-length cDNAs defined in the specification. Where a primer set	
CC	XX	comprises: (a) an oligo-dT primer and an oligonucleotide complementary	
CC	XX	to the complementary strand of a polynucleotide which comprises one of	
CC	XX	the 5602 nucleotide sequences defined in the specification, where the	
CC	XX	oligonucleotide comprises at least 15 nucleotides; or (b) a combination	
CC	XX	of an oligonucleotide comprising a sequence complementary to the	
CC	XX	complementary strand of a polynucleotide which comprises a 5'-end	
CC	XX	sequence and an oligonucleotide comprising a sequence complementary to a	
CC	XX	polynucleotide which comprises a 3'-end sequence, where the	
CC	XX	oligonucleotide comprises at least 15 nucleotides and the combination of	
CC	XX	the 5'-end sequence/3'-end sequence is selected from those defined in	
CC	XX	the specification. The primer sets can be used in antisense therapy and	
CC	XX	in gene therapy. The primers are useful for synthesizing polynucleotides,	
CC	XX	particularly full-length cDNAs. The primers are also useful for the	
CC	XX	detection and/or diagnosis of the abnormality of the proteins encoded by	

CC	the full-length cDNAs. The primers allow obtaining of the full-length
CC	cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC	AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC	AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC	represent oligonucleotides, all of which are used in the exemplification
XX	of the present invention.
XX	
SQ	Sequence 474 BP; 115 A; 134 C; 114 G; 104 T; 7 other;
	Query Match 36.0%; Score 28.8; DB 22; Length 474;
	Best Local Similarity 60.0%; Pred. No. 2.9;
	Matches 48; Conservative 0; Mismatches 32; Indels 0; Gaps
OY	1 attacatccaccagcatgagggtctgttactagtcttgctgtgcctcgcaatgg 60
Dd	206 AATCAATAATGACACAGCATGTGCAGACTCGTCTTCATGCATGCTGGTTTGTCACACTGCCGCC 147
OY	61 ggagcagctccaacttggg 80
Dd	146 GGAAATCCGACCCTGTGGGG 127
RESULT	8
ID	AACT9853
XX	AACT9853 standard; cDNA; 573 BP.
XX	
XX	AACT9853;
XX	
DT	09-FEB-2001 (first entry)
XX	
DE	Human secreted protein encoding cDNA for gene 5.
XX	
KW	Human; secreted protein; cytosolic; antiarthritic; antiasthmatic;
KW	immunosuppressive; antiarteriosclerotic; antiinflammatory; neutrotropic;
KW	neuroprotective; antidiabetic; tranquiliser; vulnerrary; antibacterial;
KW	antiporiatic; antiarrhythmic; antirheumatic; cardiant; anti-HIV;
KW	autoimmune disorder; allergic condition; cardiovascular disorder;
KW	cancer; neurological disease; tissue repair; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200055176-A2.
PN	
PD	21-SEP-2000.
XX	
PE	09-MAR-2000; 2000MO-US06057.
XX	
PR	12-MAR-1999; 99US-0124142.
PR	11-JUN-1999; 99US-0138597.
PR	03-DEC-1999; 99US-0168666.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
R1	
DR	Rosen CA, Ruben SM, Komatsoulis G;
DR	WPJ; 2000-638176/61.
DR	p-PADB; AAB44834.
XX	
PT	Novel 49 human secreted proteins useful for diagnosis, prevention and
PT	treatment of disorders including neurological, cell proliferative,
PT	cardiovascular, and autoimmune/inflammatory disorders and microbial
PT	infections
XX	
PS	Claim 1a; Page 338-339; 405pp; English.
XX	
CC	This invention describes a novel isolated polypeptide (I) comprising an
CC	amino acid sequence at least 95 % identical to a polypeptide sequence
CC	selected from 49 polypeptides encoded by polynucleotide sequences
CC	included in American Type Culture Collection (ATCC) deposit number
CC	203917, defined in the specification. The products of the invention have
CC	cytostatic, antiarthritic, antistematic, immunosuppressive, neutrotopic;
CC	antiarteriosclerotic, antiinflammatory, neuroprotective, anti diabetic;
CC	

PS Claim 8; SEQ ID 17510; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification, where a primer set
CC comprises: (a) an oligo-dt primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 4214 BP; 779 A; 1120 C; 1278 G; 1037 T; 0 other:

Query Match 36.0%; Score 28.8; DB 22; Length 4214;
Best Local Similarity 60.0%; Pred. No. 4.5;
Matches 48; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 attcacatccaccagcagcagtgagtgctgtactgctgtctgtggtcgcagctg 60
DB 4009 aatcaatgacacgacgacatgagcgtgtcctcagctgctgtgtcactgcgtcct 4068
QY 61 gggaccagtcacacttg99g 80
DB 4069 ggaatccgacgcgtgtgtg99g 4088

RESULT 11
ABA20474
ID ABA20474 standard; DNA; 6746 BP.
XX
AC ABA20474;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 12805.
XX
KW Human; noctropic; neuroprotective; cytoskeletal; dermatological; virucide;
KW immunosuppressive; antitumorigenic; anti-HIV; antibacterial; vulnary;
KW antiparkinsonian; antisticking; antidiabetic; cancer;
KW antileukemic; hepatotropic; cerebroprotective; antineoplastic;
KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225457.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 20-OCT-2000; 2000US-0242221.
 PR 01-NOV-2000; 2000US-0244611.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249246.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 01-DEC-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251836.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-02539678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI: 2001-541565/60.
 XX
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 PS Disclosure: SEQ ID NO 12805; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (AB11004-AB21534) and proteins
 CC (AB11678-AB18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are

CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases such as viral, bacterial, fungal
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 6746 BP; 1235 A; 1881 C; 2017 G; 1613 T; 0 other:

 Query Match 36.0%; Score 28.8; DB 22; Length 6746;
 Best Local Similarity 60.0%; Pred. No. 4.9;
 Matches 48; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

 QY 1 atcacatccaccagccatgagggtgtctgtactagctctgtgtgctgcagctgg 60
 Db 6513 aatcaaatgcacagcatggaagctgttccatgcactggtttgtcactgcgctct 6572
 QY 61 ggagccagctccacttggg 80
 Db 6573 ggaatccagcgtgttggg 6592

 RESULT 12
 ABA20475
 ID ABA20475 standard; DNA; 7720 BP.
 XX
 AC ABA20475;
 XX
 DT 23-JAN-2002 (first entry)
 XX
 DE Human nervous system related polynucleotide SEQ ID NO 12806.
 XX
 KW Human; nocotropic; neuroprotective; cytosolic; dermatological; virocidic;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
 KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
 KW antineumatic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antilucer; anticonvulsant; antilungal;
 KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO200159063-A2.
 XX
 PD 16-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01334.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUL-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226868.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232089.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235835.
 PR 27-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246509.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
 Rosen CA, Barash SC, Ruben SM;
 WPI: 2001-541565/60.

Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 useful for preventing, diagnosing and/or treating nervous system
 cancers and metastases -

Disclosure: SEQ ID NO 12806; 1701pp + Sequence Listing; English.

The invention relates to novel genes (ABA11004-ABA21534) and proteins (ABA1678-ABA18001) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemias;

(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal and parasitic infections.
 Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Sequence 7720 BP: 1447 A; 2142 C; 2279 G; 1852 T; 0 other;

Query Match 36.0%; Score 28.8; DB 22; Length 7720;
 Best Local Similarity 60.0%; Pred. No. 5.1;
 Matches 48; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 1 attcacatccaccagcatgaggtgctgtactagctctgtgtggtctcgcagtg 60
 ||||| ||| || ||||| || ||||| ||||| |||||
 Db 7487 aatcaaatgcacgcagcatgagctgtctgcacatgcactgtttgtcactgagctcct 7546
 ||||| ||| |||||
 OY 61 gggaccagtcacactggg 80
 ||||| |||||
 Db 7547 ggaatcgacgcgtgtgtgg 7566

RESULT 13

AAT03855
 ID AAT03855 standard; DNA: 10807 BP.

AC AAT03855;

DT 26-OCT-1996 (first entry)

DE Sheep beta-lactoglobulin coding sequence.

KM Sheep beta-lactoglobulin; signal peptide; promoter; fusion protein;
 KW human fibrinogen A-alpha chain; B-beta chain; gamma chain;
 KM transgenic animal milk; ds.

OS Ovis aries.

PN WO9523868-A1.

PD 08-SEP-1995.

PF 01-MAR-1995; 95WO-US02648.

PR 03-MAR-1994; 94US-0206176.

PA (PHAR-) PHARM PROTEINS LTD.

PA (ZYMO) ZYMOGENETICS INC.

PI Dalrymple MA, Foster DC, Garner I, Prunkard DE;

DR WPI: 1995-320582/A1.

PT Production of fibrinogen in transgenic mammals - by introducing DNA
 PT segments into the germ line of a non-human mammal and collecting
 PT milk from female progeny.

PS Disclosure; Page 69-77; 99pp; English.

This sequence encodes sheep beta-lactoglobulin and contains a unique EcoRV site at nucleotide 4245 in the 5' untranslated region of the gene. This site allows insertion of additional DNA sequences under the control of the beta-lactoglobulin promoter 3' to the transcription initiation site. The promoter and signal peptide may be used for mammal tissue-specific gene expression and secretion of human fibrinogen A-chain (see AAT03853), B-beta chain (AAT03852) and gamma chain (AAT03854) in non-human transgenic animals. A region of at least the proximal 406 bp of 5' flanking sequences (nucleotides 344-4257) is generally used, although larger sequences (nucleotides 344-4257) are preferred. Furthermore, the region surrounding the initiation ATG of one or more of the human fibrinogen sequences may be replaced with the corresponding

CC beta-lactoglobulin sequences providing a putative tissue-specific
 CC initiation environment to enhance expression.

Sequence 10807 BP: 2215 A; 3065 C; 2985 G; 2542 T; 0 other;

Query Match 35.5%; Score 28.4; DB 16; Length 10807;
 Best Local Similarity 62.9%; Pred. No. 7.4;
 Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

OY 9 ccaccagcatagagggtctgtactagctctgtgtggtctcgcagtgaggaccag 68
 ||||| ||| || ||||| || ||||| ||||| |||||
 Db 4249 cctgcagcatgaagtgctctgtctgtccttggtggcctgcctgcctgtggcgctcag 4308
 ||||| ||| |||
 OY 69 tcccaactgg 78
 ||||| |||
 Db 4309 gccatcacg 4318

RESULT 14

AAT9725
 ID AAT9725 standard; DNA: 10807 BP.

AC AAT9725;

DT 11-DEC-1997 (first entry)

DE Ovine beta-lactoglobulin gene.

KM Protein C; transgenic animal; sheep; rabbit; cattle; goat; milk;
 KW blood clotting; anticoagulant; beta-lactoglobulin; promoter; ss.

OS Ovis aries.

FN Key Location/Qualifiers

FT 5'UTR 1..4257

PN WO9720043-A1.

PD 05-JUN-1997.

PF 26-NOV-1996; 96WO-US18866.

PR 13-JUN-1996; 96US-0019692.

PR 30-NOV-1995; 95US-0565074.

PA (PPLT-) PPL THERAPEUTICS.

PA (ZYMO) ZYMOGENETICS INC.

PI Cottingham I, Foster DC, Garner I, Prunkard DE;

DR WPI: 1997-310599/28.

PT Production of protein C in transgenic animal - useful for high
 PT quantity protein C production with therapeutic value

PS Disclosure; Page 66-74; 99pp; English.

This DNA sequence comprises the ovine beta-lactoglobulin gene including the promoter region. A DNA segment encompassing the 5' flanking promoter region can be utilized in novel constructs for the expression of human protein C in the milk of a transgenic animal. A claimed method involves: (a) providing a DNA construct comprising DNA encoding a secretion signal and a protein C propeptide, operably linked to DNA encoding two-chain cleavage site-modified protein C (see AAT25085-86), the 2 DNA sequences being linked to elements required for protein C expression in a mammary gland of a host female animal, such as beta-lactoglobulin gene promoter; and (b) using the DNA construct to breed a transgenic animal (esp. sheep, rabbit, cattle, goat) that produces protein C in its milk, at least 90% of the protein C being in the two-chain

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2002, 23:55:49 ; Search time 77.96 Seconds
(without alignments)
252,061 Million cell updates/sec

Title: US-09-426-776a-11

Perfect score: 80

Sequence: 1 attcaccatccaccagcatg.....gggaccagctccacttggg 80

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_NA :
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	28.4	35.5	10807	1 US-08-206-176-7	Sequence 7, Appl
2	28.4	35.5	10807	2 US-08-756-506-5	Sequence 5, Appl
3	27.6	34.5	824	1 US-08-154-019-29	Sequence 29, Appl
4	27.6	34.5	824	1 US-08-461-333-29	Sequence 29, Appl
5	27.6	34.5	824	3 US-08-464-167-29	Sequence 29, Appl
6	27.6	34.5	824	3 US-09-158-313-29	Sequence 29, Appl
7	27.6	34.5	824	3 US-08-476-798-29	Sequence 29, Appl
8	25.4	31.8	1607	3 US-08-753-007A-3	Sequence 3, Appl
9	25.4	31.8	1607	3 US-09-398-496-3	Sequence 3, Appl
10	25.4	31.8	2467	3 US-08-753-007A-1	Sequence 1, Appl
11	25.4	31.8	2467	3 US-09-398-196-1	Sequence 1, Appl
12	25	31.2	987	4 US-09-159-106-12	Sequence 12, Appl
13	25	31.2	1177	4 US-09-159-106-1	Sequence 1, Appl
14	25	31.2	1516	4 US-09-159-106-10	Sequence 10, Appl
15	24.4	30.5	4403765	4 US-09-103-840A-2	Sequence 2, Appl
16	24.4	30.5	4411529	4 US-09-103-840A-1	Sequence 1, Appl
17	24.2	30.2	2296	2 US-08-899-336-2	Sequence 3, Appl
18	24.2	30.2	33529	2 US-09-144-085-3	Sequence 3, Appl
19	23.8	29.8	2018	2 US-08-557-973-1	Sequence 1, Appl
20	23.8	29.8	35081	2 US-08-752-760A-1	Sequence 1, Appl
21	23.6	29.5	274	3 US-08-444-818-63	Sequence 63, Appl
22	23.6	29.5	1788	2 US-08-722-806A-1	Sequence 1, Appl
23	23.6	29.5	3495	1 US-09-337-028-1	Sequence 1, Appl
24	23.6	29.5	3495	1 US-08-446-038B-2	Sequence 2, Appl
25	23.6	29.5	3495	1 US-08-446-010B-2	Sequence 2, Appl
26	23.6	29.5	3495	1 US-08-805-445-2	Sequence 2, Appl
27	23.6	29.5	3495	2 US-08-064-067D-2	Sequence 2, Appl

ALIGNMENTS

28	23.6	29.5	3495	2 US-09-066-208-2	Sequence 2, Appl
29	23.6	29.5	3495	4 US-08-980-080-3	Sequence 3, Appl
30	23.6	29.5	3629	1 US-08-097-997A-8	Sequence 8, Appl
31	23.6	29.5	3629	3 US-08-665-574C-8	Sequence 8, Appl
32	23.6	29.5	3629	4 US-08-946-994-8	Sequence 8, Appl
33	23.6	29.5	6785	3 US-08-444-818-65	Sequence 65, Appl
34	23.6	29.5	7310	3 US-08-444-818-74	Sequence 74, Appl
35	23.6	29.5	8316	3 US-08-444-818-88	Sequence 88, Appl
36	23.6	29.5	8987	3 US-08-444-818-137	Sequence 137, App
37	23.6	29.5	9185	3 US-08-444-818-122	Sequence 122, App
38	23.6	29.5	9185	3 US-08-444-818-123	Sequence 123, App
39	23.6	29.5	9379	4 US-08-444-818-176	Sequence 176, App
40	23.6	29.5	9379	4 US-09-388-874-1	Sequence 1, Appl
41	23.6	29.5	9401	1 US-07-910-760-9	Sequence 9, Appl
42	23.6	29.5	9401	1 US-08-440-519-9	Sequence 9, Appl
43	23.6	29.5	9401	4 US-08-440-549-9	Sequence 9, Appl
44	23.6	29.5	9401	5 PCT-US91-02225-9	Sequence 9, Appl
45	23.4	29.2	1195	4 US-08-943-731-152	Sequence 152, App

RESULT 1

US-08-206-176-7
Sequence 7, Application US/08206176
Patent No. 5639940
GENERAL INFORMATION:
APPLICANT: Garner, Ian
APPLICANT: Dalrymple, Michael A
APPLICANT: Prunkard, Donna E
APPLICANT: Foster, Donald C
TITLE OF INVENTION: Production of Fibrinogen in Transgenic
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Zymogenetics, Inc.
STREET: 4225 Roosevelt Way, N.E.
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/206,176
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-648
REFERENCE NUMBER: 93-15
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-547-8080 ext 322
TELEFAX: 206-548-2329
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 10807 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: ovine beta-lactoglobulin
US-08-206-176-7

Query Match 35.5%; Score 28.4; DB 1; Length 10807;
Best Local Similarity 62.9%; Pred. No. 0.87;
Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

Patent No. 6013857

GENERAL INFORMATION:

APPLICANT: Deboer, Herman A.

APPLICANT: Strijker, Rein

APPLICANT: Heyneker, Herbert L.

APPLICANT: Platenburg, Gerald

APPLICANT: Lee, Sang He

APPLICANT: Pieper, Frank

APPLICANT: Krimpenfort, Paul J.A.

TITLE OF INVENTION: Production of Recombinant Polypeptides

TITLE OF INVENTION: by Bovine Species and Transgenic Methods

NUMBER OF SEQUENCES: 38

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew

STREET: One Market Plaza, Stewart Tower, Suite 2000

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/464,167

FILING DATE: 05-JUN-1995

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/077,788

FILING DATE: 15-JUN-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/895,956

FILING DATE: 15-JUN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/619,131

FILING DATE: 27-NOV-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/444,745

FILING DATE: 01-DEC-1989

ATTORNEY/AGENT INFORMATION:

NAME: Liebescheutz, Joe O.

REGISTRATION NUMBER: 37,505

REFERENCE/DOCKET NUMBER: 16994-003124

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415-543-9600

TELEFAX: 415-543-5043

INFORMATION FOR SEQ ID NO: 29:

SEQUENCE CHARACTERISTICS:

LENGTH: 824 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-464-167-29

[illegible]

```
; Patent No. 6066725
; GENERAL INFORMATION:
; APPLICANT: Deboer, Herman A.
; APPLICANT: Strijker, Rein
; APPLICANT: Heyneker, Herbert L.
; APPLICANT: Platenburg, Gerald
; APPLICANT: Lee, Sang He
; APPLICANT: Pieper, Frank
; APPLICANT: Krumpenfort, Paul J.A.
; TITLE OF INVENTION: Production of Recombinant Polypeptides
; TITLE OF INVENTION: by Bovine Species and Transgenic Methods
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/158,313
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,798
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/895,956
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/619,131
; FILING DATE: 27-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/444,745
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebeschuetz, Joe O.
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 16994-003125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 824 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-158-313-29

Query Match          34.5%; Score 27.6; DB 3; Length 824;
Best Local Similarity 63.6%; Pred. No. 0.89;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
```

```
RESULT 7
US-08-476-798-29
; Sequence 29, Application US/08476798
```

```
; Patent No. 6140552
; GENERAL INFORMATION:
; APPLICANT: Deboer, Herman A.
; APPLICANT: Strijker, Rein
; APPLICANT: Heyneker, Herbert L.
; APPLICANT: Platenburg, Gerald
; APPLICANT: Lee, Sang He
; APPLICANT: Pieper, Frank
; APPLICANT: Krumpenfort, Paul J.A.
; TITLE OF INVENTION: Production of Recombinant Polypeptides
; TITLE OF INVENTION: by Bovine Species and Transgenic Methods
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,798
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/077,788
; FILING DATE: 15-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/895,956
; FILING DATE: 15-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/619,131
; FILING DATE: 27-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/444,745
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Liebeschuetz, Joe O.
; REGISTRATION NUMBER: 37,505
; REFERENCE/DOCKET NUMBER: 16994-003125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-543-9600
; TELEFAX: 415-543-5043
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 824 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-476-798-29

Query Match          34.5%; Score 27.6; DB 3; Length 824;
Best Local Similarity 63.6%; Pred. No. 0.89;
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;
```

```
RESULT 8
US-08-753-007A-3/C
; Sequence 3, Application US/08753007A
```

```

? STREET: 225 Franklin Street
? CITY: Boston
? STATE: MA
? COUNTRY: US
? ZIP: 02110-2804
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? OPERATING SYSTEM: DOS
? SOFTWARE: FASTSEQ Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/398,496
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/753,007
? FILING DATE: 19-NOV-1996
? APPLICATION NUMBER: 08/699,591
? FILING DATE: 19-AUG-1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Fasse, J. Peter
? REGISTRATION NUMBER: 32,983
? REFERENCE/DOCKET NUMBER: 07334/022001
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 617-542-5070
? TELEFAX: 617-542-8906
? TELEX:
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1607 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: Coding Sequence
? LOCATION: 79...621
? OTHER INFORMATION:
?
US-09-398-496-3

Query Match      31.88; Score 25.4; DB 3; Length 1607;
Best Local Similarity 61.28; Pred. No. 5.9;
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY      8 tccaccacacttgagggtgctgtacctagccttgcgtgtgcttcggagggaggaacca 67
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db       768 TCTGCGCAGATGATGATGCATCTCCCTCTCTGTTTTGGCTTGACGTAAGCGCACCA 709

QY      68 gtccaac 74
        |  ||
Db       708 CACGAGAC 702

RESULT 10
US-08-753-007A-1/c
? Sequence 1, Application US/08753007A
? Patent No. 6074841
? GENERAL INFORMATION:
? APPLICANT: Gearling, David P.
? TITLE OF INVENTION: DON-I GENE AND POLYPEPTIDES
? NUMBER OF SEQUENCES: 33
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Fish & Richardson P.C.
? STREET: 225 Franklin Street
? CITY: Boston
? STATE: MA
? COUNTRY: US
? ZIP: 02110-2804
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
```

```

, FILING DATE: 19-NOV-1996
, APPLICATION NUMBER: 08/699,591
, FILING DATE: 19-AUG-1996
, ATTORNEY/AGENT INFORMATION:
, NAME: Fasse, J. Peter
, REGISTRATION NUMBER: 32,983
, REFERENCE/DOCKET NUMBER: 07334/022001
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: 617-542-5070
, TELEFAX: 617-542-8906
, TELEX:
, INFORMATION FOR SEQ ID NO: 1:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 2467 base pairs
, TYPE: nucleic acid
, STRANDEDNESS: single
, TOPOLOGY: circular
, MOLECULE TYPE: cDNA
, FEATURE:
, NAME/KEY: Coding Sequence
, LOCATION: 79...1893
, OTHER INFORMATION:
, US-09-398-496-1

```

Query Match	31.88;	Score 25.4;	DB 3;	Length 2467;
Best Local Similarly	61.28;	Pred. No. 6.5;		
Matches 41; Conservative	0;	Mismatches 26;	Indels 0;	Gaps 0;

QY 8 tccaccagcatgagggtgcctgtactacgtcctgtgctctcgcaagtgggagcca 67
|| || | ||| | ||| | ||| | ||| | ||| | ||| | |||
Db 685 TCtCCCGGAGATGATGCATCTCCTCCTCTCTTTTGGTCCTTGCAAGGCCACCA 626

QY	68	gtccaac	74
Db	625	CACAGAC	619

```

RESULT 12
US-09-159-106-12/c
; Sequence 12, Application US/09159106
; Patent No. 6284509
; GENERAL INFORMATION:
; APPLICANT: Ferrer, Pau
; APPLICANT: Diers, Ivan
; APPLICANT: Halkier, Torben
; APPLICANT: Hedegaard, Lisbeth
; TITLE OF INVENTION: An Enzyme With -1,3-Glucanase
; TITLE OF INVENTION: Activity
; FILE REFERENCE: 4693.204-US
; CURRENT APPLICATION NUMBER: US/09/159,106
; CURRENT FILING DATE: 1998-09-23
; EARLIER APPLICATION NUMBER: 0427/96
; EARLIER FILING DATE: 1996-12-04
; EARLIER APPLICATION NUMBER: 0885/96
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: PCT/DK97/00160
; EARLIER FILING DATE: 1997-04-14
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Oerskovia xanthineolytica
US-09-159-106-12

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 15:00:50 ; Search time 14.76 Seconds
(without alignments)
136.712 Million cell updates/sec

Title: US-09-426-776A-10

Perfect score: 95

Sequence: 1 MRVLVALAVAGDQSNL 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	93.7	1788	2 T31095	vitellogenin II precursor - mummichog
2	58	61.1	1687	2 T43144	vitellogenin II precursor - mummichog
3	47	49.5	258	2 JH0472	apolipoprotein E - mummichog
4	45	47.4	678	2 S46963	exed protein - Aet
5	44	46.3	180	1 LGCT	beta-lactoglobulin
6	44	46.3	180	1 LGCT	beta-lactoglobulin
7	44	46.3	180	1 LGCT	beta-lactoglobulin
8	44	46.3	180	1 LGCT	beta-lactoglobulin
9	44	46.3	180	1 LGCT	beta-lactoglobulin
10	44	46.3	180	1 LGCT	beta-lactoglobulin
11	43.5	45.8	222	2 JH0472	apolipoprotein E - mummichog
12	43	45.3	127	2 B75353	beta-casein precursor - mummichog
13	43	45.3	127	2 B75353	beta-casein precursor - mummichog
14	43	45.3	127	2 B75353	beta-casein precursor - mummichog
15	43	45.3	127	2 B75353	beta-casein precursor - mummichog
16	42	44.2	174	2 A64442	pheromone-binding protein - mummichog
17	42	44.2	174	2 A64442	pheromone-binding protein - mummichog
18	42	44.2	174	2 A64442	pheromone-binding protein - mummichog
19	42	44.2	174	2 A64442	pheromone-binding protein - mummichog
20	42	44.2	174	2 A64442	pheromone-binding protein - mummichog
21	42	44.2	174	2 A64442	pheromone-binding protein - mummichog
22	42	44.2	174	2 A64442	pheromone-binding protein - mummichog
23	41.5	43.7	222	2 JH0472	apolipoprotein E - mummichog
24	41	43.2	109	2 B75353	beta-casein precursor - mummichog
25	41	43.2	109	2 B75353	beta-casein precursor - mummichog
26	41	43.2	109	2 B75353	beta-casein precursor - mummichog
27	41	43.2	109	2 B75353	beta-casein precursor - mummichog
28	41	43.2	109	2 B75353	beta-casein precursor - mummichog
29	41	43.2	109	2 B75353	beta-casein precursor - mummichog

30	41	43.2	312	2 S18387	heme oxygenase - mummichog
31	41	43.2	316	2 S25478	apolipoprotein E - mummichog
32	41	43.2	316	2 T45996	apolipoprotein E - mummichog
33	41	43.2	481	2 T23729	apolipoprotein E - mummichog
34	41	43.2	597	2 C69283	hypothetical protein - mummichog
35	41	43.2	821	2 B84509	hypothetical protein - mummichog
36	41	43.2	896	2 AFI409	probable Na/H antiporter - mummichog
37	41	43.2	1173	2 H83189	the two components of the Na/H antiporter - mummichog
38	40	42.1	111	2 T18094	DNA polymerase III - mummichog
39	40	42.1	185	2 A49713	hypothetical protein - mummichog
40	40	42.1	264	2 A75354	uroplakin II precursor - mummichog
41	40	42.1	266	1 MOECMP	hypothetical protein - mummichog
42	40	42.1	266	2 H90944	phosphotransferase - mummichog
43	40	42.1	266	2 C85793	mannose-specific P-1-ATase - mummichog
44	40	42.1	300	2 D87497	PTS enzyme IIC, mannanase - mummichog
45	40	42.1	312	2 B83749	sugar transport system - mummichog

ALIGNMENTS

RESULT 1
T31095
vitellogenin II precursor - Oreochromis aureus
C:Species: Oreochromis aureus
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C:Accession: T31095
R:Lim, E.H.; Lam, T.J.; Ding, J.L.
Submitted to the EMBL Data Library, August 1997
A:Description: Cloning of full-length Oreochromis aureus vitellogenin cDNA and its
A:Reference number: 220978
A:Accession: T31095
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-1788 <LIM>
A:Cross-references: EMBL:AF017250; NID:94102880; PID:94102881; PIDN:AA001615.1
C:Superfamily: vitellogenin

Query Match 93.7%; Score 89; DB 2; Length 1788;
Best Local Similarity 100.0%; Pred. No. 4.3e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRVLVALAVAGDQSNL 20
DB 1 MRVLVALAVAGDQSNL 20

RESULT 2
T43144
vitellogenin II precursor - mummichog
C:Species: Fundulus heteroclitus (mummichog)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 18-Feb-2000
C:Accession: T43144
R:Laurent, G.J.; Hoch, K.L.; Denslow, N.; Byrne, B.M.; Wallace, R.A.
Submitted to the EMBL Data Library, September 1996
A:Description: Derivation of oocyte and egg proteins from parental vitellogenins II
D:Reference number: 222316
A:Accession: T43144
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1687 <LAF>
A:Cross-references: EMBL:U70826; NID:91621358; PID:91621359; PIDN:AA017152.1
C:Experimental source: estrogen-induced liver
C:Superfamily: vitellogenin
C:Keywords: egg yolk; phosphoprotein
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-1687/Product: vitellogenin II #status predicted <MAT>

Query Match 61.1%; Score 58; DB 2; Length 1687;
Best Local Similarity 76.5%; Pred. No. 1.5;

Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 MRVLTALVALAVGDS 17
 DB 1 MKRVLTALVALAVGDS 17

RESULT 3

apolipoprotein A-I precursor - Atlantic salmon

C:Species: Salmo salar (Atlantic salmon)

C:Date: 31-Mar-1992 #sequence, revision 31-Mar-1992 #text_change 17-Mar-2000

C:Accession: JH0472; S26810; S18833

A:Title: The salmon gene encoding apolipoprotein A-I: cDNA sequence, tissue expression

A:Reference number: JH0472; M01D:92009208

A:Accession: JH0472

A:Molecule type: mRNA

A:Residues: 1-258 <POM>

A:Cross-references: GB:X52237

A:Experimental source: liver

A:Note: the authors translated the codon ACG for residue 252 as Ser

R.Powell, R.

submitted to the EMBL Data Library, March 1990

A:Reference number: S26810

A:Accession: S26810

A:Molecule type: preliminary

A:Residues: 1-123, 'T', 125-143, 'Q', 145-160, 'D', 162-258 <PO2>

A:Cross-references: EMBL:X52237; NID:964355; PID:964356

C:Comment: Apolipoproteins are the major agents for transporting cholesterol and fatty

C:Superfamily: apolipoprotein A-I

C:Keywords: lipid binding; lipoprotein

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-31/Domain: propeptide #status predicted <PRE>

F:32-258/Product: apolipoprotein A-I #status predicted <PRO>

Query Match 49.5%; Score 47; DB 2; Length 258;
 Best Local Similarity 55.6%; Pred. No. 13;
 Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 MRVLTALVALAVGDS 18
 DB 1 MKRVLTALVALAVGDS 18

RESULT 4

exed protein - Aeromonas salmonicida

C:Species: Aeromonas salmonicida

C:Date: 06-Jan-1995 #sequence, revision 06-Jan-1995 #text_change 08-Oct-1999

C:Accession: J39678; S46963

R:Karilyshev, A.V.; MacIntyre, S.

Gene 158, 77-82, 1995

A:Title: Cloning and study of the genetic organization of the exx gene cluster of Aerom

A:Reference number: J39675; M01D:95309729

A:Accession: J39678

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-678 <RES>

A:Cross-references: EMBL:X80505; NID:9516349; PIDN:CA56668.1; PID:9516351

C:Genetics:

A:Gene: exed

Query Match 47.4%; Score 45; DB 2; Length 678;
 Best Local Similarity 55.0%; Pred. No. 58;
 Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 2 RVVLTALVALAVGDS 21
 DB 1 MKRVLTALVALAVGDS 21

DB 352 QVLVEAIVETADGGLMG 371

RESULT 5

beta-lactoglobulin precursor - goat

C:Species: Capra aegagrus hircus (domestic goat)

C:Date: 17-May-1985 #sequence, revision 12-Apr-1996 #text_change 22-Jun-1999

C:Accession: A03220; S14507; S42800; S42801

R:Preaux, G.; Braunitzer, G.; Schrank, B.; Stangl, A.

Hoppe-Seyler's Z. Physiol. Chem. 360, 1595-1604, 1979

A:Title: The amino acid sequence of goat beta-lactoglobulin.

A:Reference number: A91682; M01D:80070611

A:Accession: A03220

A:Molecule type: protein

A:Residues: 19-180 <PRE>

R:Polch, J.M.; Coll, A.; Sanchez, A.

submitted to the EMBL Data Library, March 1991

A:Reference number: S14507

A:Accession: S14507

A:Molecule type: mRNA

A:Residues: 1-180 <KIM>

A:Cross-references: EMBL:Z19569; NID:9437751; PIDN:CA79623.1; PID:9437752

A:Accession: S42800

A:Molecule type: mRNA

A:Residues: 1-32 <K12>

A:Cross-references: EMBL:Z19570; NID:9437753; PIDN:CA79624.1; PID:9437754

C:Comment: Under physiological conditions beta-lactoglobulin exists as an equilibrium

C:Superfamily: lipocalin; lipocalin homology

C:Keywords: milk

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-180/Product: beta-lactoglobulin #status predicted <MAM>

F:28-178/Domain: lipocalin homology <LIP>

F:84-178,124-137/Disulfide bonds: #status predicted

Query Match 46.3%; Score 44; DB 1; Length 180;
 Best Local Similarity 55.6%; Pred. No. 27;
 Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 MRVLTALVALAVGDS 18
 DB 1 MKVLTALVALAVGDS 18

RESULT 6

beta-lactoglobulin precursor - sheep

N:Alternate names: beta-lactoglobulin A; beta-lactoglobulin B; beta-lactoglobulin C;

C:Species: Ovis orientalis aries; Ovis ammon aries (domestic sheep)

C:Date: 17-May-1985 #sequence, revision 19-Apr-1996 #text_change 22-Jun-1999

C:Accession: J00749; J00748; A30011; B30011; S02136; A25136; A03221; S04955

R:Ali, S.; McClenaghan, M.; Simons, J.P.; Clark, A.J.

Gene 91, 201-207, 1990

A:Title: Characterisation of the alleles encoding ovine beta-lactoglobulins A and B.

A:Reference number: J00748; M01D:91007276

A:Accession: J00749

A:Molecule type: DNA

A:Residues: 1-180 <ALIB>

A:Cross-references: GB:M32232

A:Experimental source: beta-lactoglobulin B

A:Accession: J00748

A:Molecule type: DNA

A:Residues: 1-37, 'Y', 39-180 <ALIA>

A:Cross-references: GB:M32232

A:Experimental source: beta-lactoglobulin A

R; Ali, S.; Clark, A.J.
J. Mol. Biol. 199, 415-426, 1988
A>Title: Characterisation of the gene encoding ovine beta-lactoglobulin. Similarity to t
A:Reference number: A92942; MUID:88172489
A:Accession: A30011
A:Molecule type: DNA
A:Residues: 1-180 <ALI>
A:Cross-references: GB:X14971
A:Experimental source: beta-lactoglobulin I
A:Accession: B30011
A:Molecule type: DNA
A:Residues: 1-37, 'Y', 39-102, 'N', 104-180 <ALI2>
A:Cross-references: GB:X07009
A:Experimental source: beta-lactoglobulin II
R; Harris, S.; Ali, S.; Anderson, S.; Archibald, A.L.; Clark, A.J.
Molec. Acids Res. 16, 10379-10380, 1988
A>Title: Complete nucleotide sequence of the genomic ovine beta-lactoglobulin gene.
A:Reference number: S02136; MUID:89057492
A:Accession: S02136
A>Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-180 <HR>
A:Cross-references: EMBL:X12817; NID:g1315; PIDN:CAA31305.1; PID:g1314
P; Gaye, P.; Hue-Delamare, D.; Mercier, J.C.; Soulier, S.; Viotte, J.L.; Furet, J.P.
Biochimie 68, 1097-1107, 1986
A>Title: Ovine beta-lactoglobulin messenger RNA: nucleotide sequence and mRNA levels dur
A:Reference number: A25136; MUID:87049827
A:Accession: A25136
A:Molecule type: mRNA
A:Residues: 1-180 <GAR>
A:Cross-references: GB:X04520; NID:g1315; PIDN:CAA28204.1; PID:g1316
R; Pireaux, G.; Braunlietzer, G.; Kolde, H.J.
Arch. Int. Physiol. Biochim. 88, B45-B46, 1980
A>Title: Primary structure of ovine beta-lactoglobulin.
A:Reference number: A03221; MUID:80219294
A:Accession: A03221
A:Molecule type: Protein
A:Residues: 19-37, 'Y', 39-180 <PRE>
R; Reinhardt, G.; Godovac-Zimmermann, J.; Contl, A.
Biol. Chem. Hoppe-Seyler 370, 757-762, 1989
A>Title: Isolation and complete primary sequence of a new ovine wild-type beta-lactoglob
A:Reference number: S04955; MUID:89374823
A:Accession: S04955
A:Molecule type: protein
A:Residues: 19-37, 'Y', 39-165, 'Q', 167-180 <ERH>
A:Experimental source: beta-lactoglobulin C
C:Comment: This protein is the major milk whey protein of ruminants and is produced in t
C:Genetics: Under physiological conditions beta-lactoglobulin exists as an equilibrium m
A:Gene: BLG
A:introns: 32/3; 79/2; 104/1; 141/1; 176/1
C:Superfamily: lipocalin; lipocalin homology
C:Keywords: milk; polymorphism
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-180/Product: beta-lactoglobulin #status experimental <MAT>
F:28-178/Domain: lipocalin homology <IRP>
F:84-178,124-137/Dissulfide bonds: #status predicted

Query Match 46.3%; Score 44; DB 1; Length 180;
Best local similarity 55.6%; Pred. No. 27;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 MRVLVLAVALAVGDOS 18
I: I:I:I:I:I:I:I:I I I:
DB 1 MRCLTALGLACGVDA 18

RESULT 7
JC6549
apolipoprotein E precursor - sheep
C:Species: Ovis sp. (sheep)
C>Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 07-May-1999

C:Accession: J06549
R:Komatsu, Y.; Horluch, M.; Ishiguro, N.; Matsui, T.; Shinaagawa, M.
Gene 208, 131-138, 1998
A:Title: Characterization of the sheep apolipoprotein E (APOE) gene and allelic variants
A:Reference number: JC6549; MIDID:98201606
A:Accession: JC6549
A:Molecule type: mRNA
A:Residues: 1-316 <KOM>
C:Genetics:
A:Gene: apoE
A:Introns: 15/1: 78/2
C:Superfamily: apolipoprotein A-I
C:Keywords: lipoprotein
F:1-18/Dominant signal sequence #status predicted <Sig>
F:19-316/Product: apolipoprotein E #status experimental <Mat>

Query Match 46.3% Score 44; DB 2; Length 316;
Best Local Similarity 50.0%; Pred. No. 43;
Matches 10; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

OY 1 MRVLALVALAVAGDQSNL 20
 |::|::|::|::|::|::|
Db 1 MKVLTVALVALLAGCAADM 20

RESULT 8
T25169
hypothetical protein T23P1.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T25169
R:Wilkinson, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19990
A:Accession: T25169
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-530 <WIL>
A:Cross-references: EMBL:T281129; PIDN:CA803405.1; GSPDB:GN00023; CESP:T23P1.6
A:Experimental source: clone T23P1
C:Genetics:
A:Gene: CESP:T23P1.6
A:Map position: 5
A:Introns: 16/3
C:Superfamily: gliadin

Query Match 46.3% Score 44; DB 2; Length 330;
Best Local Similarity 45.0%; Pred. No. 45;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 1 MRVLALVALAVAGDQSNL 20
 ||:::|||||||::|
Db 1 MRCVITLFAVALVAQAASSI 20

RESULT 9
S06958
sphingomyelin phosphodiesterase, acidic, inactive splice form 2 - human
N:Alternate names: acid sphingomyelinase, splice form 2
C:Species: Homo sapiens (man)
C>Date: 22-Jan-1993 #sequence_revision 24-Nov-1999 #text_change 21-Jan-2000
C:Accession: S06958; A39825
R:Quiltern, L.E.; Schuchman, E.H.; Levran, O.; Sucht, M.; Ferlinz, K.; Relinke, H.;
EMMO J. 8, 2469-2473, 1989
A:Title: Isolation of cDNA clones encoding human acid sphingomyelinase: occurrence
A:Reference number: S06957; MIDID:90060003
A:Accession: S06958
A:Molecule type: mRNA
A:Residues: 270-585 <QUIL>
A:Cross-references: EMBL:X52679; NID:g28881; PIDN:CAA36902.1; PID:g825629
R:Schuchman, E.H.; Sucht, M.; Takahashi, T.; Sandhoff, K.; Desnick, R.J.

J. Biol. Chem. 266, 8531-8539, 1991
 A:Title: Human acid sphingomyelinase. Isolation, nucleotide sequence, and expression of
 A:Reference number: A39825; MUID:91217097
 A:Accession: A39825
 A:Molecule type: mRNA
 A:Residues: 1-362, 'IGGFYALSPYRGLRLISLNMFCSRENFLL', 'INSTDPAGLOQLVGELOAAEDKGD', 375-565
 A:Cross-references: GB:M59916; NID:9179094; PIDN:AAAS8377.1; PID:9179095
 A:Comment: This form lacked catalytic activity when expressed in COS-1 cells. Its function
 C:Comment: For the active splice form 1, see PIR:A39825.
 C:Genetics:
 A:Gene: GDB:SMPD1
 A:Cross-references: GDB:128144; OMIM:257200
 A:Map position: 11p15.4-11p15.4
 A:Introns: 104/3; 362/2; 375/3; 401/2; 450/1
 C:Superfamily: acid sphingomyelinase; phosphoesterase core homology; saposin repeat homol
 C:Keywords: alternative splicing
 F:81-115/Domain: saposin repeat homology <SNP>
 F:200-320/Domain: phosphoesterase core homology <PEC>

Query Match 46.3%; Score 44; DB 2; Length 585;
 Best Local Similarity 71.4%; Pred. No. 72;
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 3 VLVALAVALAVCD 16
 |||||:||||:
 Db 36 VLALALALALALSD 49

RESULT 10
 sphingomyelin phosphodiesterase (EC 3.1.4.12), acidic, splice form 1 precursor - human
 A39825
 N:Alternate names: acid sphingomyelinase
 C:Species: Homo sapiens (man)
 C:Date: 20-Mar-1992 #sequence, revision 07-Jul-1995 #text, change 24-Nov-1999
 C:Accession: S27009; S36357; A39825; JX0276; S06957; I55567; A42801; B42801; D42
 C:Newzeila, D.; Stoffel, W.
 Biol. Chem. Hoppe-Seyler 377, 1233-1238, 1992
 A:Title: Molecular cloning of the acid sphingomyelinase of the mouse and the organization
 A:Reference number: S27009; MUID:93183402
 A:Accession: S27009
 A:Molecule type: DNA
 A:Residues: 1-629 <NEW>
 A:Cross-references: EMBL:X63600; NID:9556808
 R: Hofmann, K.
 submitted to the EMBL Data Library, December 1991
 A:Reference number: S36357
 A:Accession: S36357
 A:Molecule type: DNA
 A:Residues: 1-321, 'T', 323-629 <HOF>
 A:Cross-references: EMBL:X63600; NID:9556808; PIDN:CAA45145.1; PID:9556809
 R:Schuchman, E.H.; Suchi, M.; Takahashi, T.; Sandhoff, K.; Desnick, R.J.
 J. Biol. Chem. 266, 8531-8539, 1991
 A:Title: Human acid sphingomyelinase. Isolation, nucleotide sequence, and expression of
 A:Reference number: A39825; MUID:91217097
 A:Accession: A39825
 A:Molecule type: mRNA
 A:Residues: 1-629 <SCH>
 A:Cross-references: GB:M59916; NID:9179094; PIDN:AAAS8377.1; PID:9179095
 A:Note: polymorphisms were demonstrated at positions 322 and 506
 R:Ida, H.; Renner, O.M.; Eto, Y.; Chan, W.Y.
 J. Biochem. 114, 15-20, 1993
 A:Title: Cloning of a human acid sphingomyelinase cDNA with a new mutation that renders
 A:Reference number: JX0276; MUID:94012573
 A:Accession: JX0276
 A:Molecule type: mRNA
 A:Residues: 1-35, 38-156, 'R', 158-321, 'T', 323-505, 'G', 507-629 <IDA>
 A:Cross-references: EMBL:X59960; NID:9402620; PIDN:CAA42584.1; PID:9402621
 A:Note: the authors translated the codon CTC for residue 85 as Thr and ACC for residue B
 R:Quintern, L.E.; Schuchman, E.H.; Levran, O.; Suchi, M.; Ferlinz, K.; Reinke, H.; Sand
 EMO J. 8, 2469-2473, 1989
 A:Title: Isolation of cDNA clones encoding human acid sphingomyelinase: occurrence of al

A:Reference number: S06957; MUID:90060003
 A:Accession: S06957
 A:Molecule type: mRNA
 A:Residues: 128-629 <QUI>
 A:Cross-references: EMBL:X52678; NID:928879; PIDN:CAA36901.1; PID:928880
 A:Note: parts of this sequence were confirmed by peptide sequencing
 R:Levan, O.; Desnick, R.O.; Schuchman, E.H.
 J. Clin. Invest. 88, 806-810, 1991
 A:Title: Niemann-Pick type B disease. Identification of a single codon deletion in t
 A:Reference number: I55567; MUID:91358737
 A:Accession: I55567
 A>Status: translation not shown; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 119-120, 'H', 122 <LEV>
 A:Cross-references: GB:S55766; NID:9234719; PIDN:AAJ19680.1; PID:9234720
 R:Takahashi, T.; Suchi, M.; Desnick, R.J.; Takada, G.; Schuchman, E.H.
 J. Biol. Chem. 267, 12552-12558, 1992
 A:Title: Identification and expression of five mutations in the human acid sphingomy
 opathic and non-neuropathic forms.
 A:Reference number: A42801; MUID:92316934
 A:Contents: annotation; characterization of mutations
 A:Note: substitution of Ile for 382-Met or Ser for 383-Asn result in complete inacti
 tion and lead to milder (type B) disease
 C:Comment: Two isoforms, neutral and acidic, have been identified. The acidic isofor
 C:Comment: For the inactive splice form 2, see PIR:S06950.
 C:Genetics:
 A:Gene: GDB:SMPD1
 A:Cross-references: GDB:128144; OMIM:257200
 A:Map position: 11p15.4-11p15.4
 A:Introns: 104/3; 362/2; 419/3; 445/2; 494/1
 A:Note: a defect in this gene may result in Niemann-Pick disease
 C:Complex: monomer
 C:Function:
 A:Description: catalyzes the hydrolysis of sphingomyelin to form phosphorylcholine an
 C:Superfamily: acid sphingomyelinase; phosphoesterase core homology; saposin repeat
 C:Keywords: acetylated amino end; alternative splicing; glycoprotein; lysosome; meta
 F:1-48/Domain: signal sequence #status predicted <SIG>
 F:49-629/Product: sphingomyelin phosphodiesterase #status predicted <MNT>
 F:81-115/Domain: saposin repeat homology <SNP>
 F:200-320/Domain: phosphoesterase core homology <PEC>
 F:49/Modified site: acetylated amino end (Asp) (in mature form) #status predicted
 F:86,175,335,395,503,520/Binding site: carbohydrate (asn) (covalent) #status predict
 F:89-165,92-157,120-131/Disulfide bonds: #status predicted
 F:251,319,439/Active site: Asp, His, Tyr #status predicted

Query Match 46.3%; Score 44; DB 1; Length 629;
 Best Local Similarity 71.4%; Pred. No. 77;
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Oy 3 VLVALAVALAVCD 16
 |||||:||||:
 Db 36 VLALALALALALSD 49

RESULT 11
 JCI1384
 beta-casein precursor - goat
 C:Species: Capra aegagrus hircus (domestic goat)
 C:Date: 10-Jun-1993 #sequence, revision 10-Jun-1993 #text, change 23-Feb-1997
 C:Accession: JCI1384
 R:Roberts, B.; Dittullio, P.; Vitale, J.; Hehr, K.; Gordon, K.
 Gene 121, 255-262, 1992
 A:Title: Cloning of the goat beta-casein-encoding gene and expression in transgenic
 A:Reference number: JCI1384; MUID:93077039
 A:Accession: JCI1384
 A:Molecule type: DNA
 A:Residues: 1-222 <ROB>
 A:Cross-references: GB:M90556
 C:Genetics:
 A:Gene: CSN2
 A:Introns: 17/3; 26/3; 35/3; 45/3; 57/3; 221/3
 C:Superfamily: beta-casein

C:Keywords: milk; phosphoprotein

Query Match 45.8%; Score 43.5; DB 2; Length 222;
Best Local Similarity 52.4%; Pred. No. 38;
Matches 11; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

OY 1 MRVLAVALAVG-DQSNL 20
1 MKVLAVALAVG-DQSNL 21

RESULT 12

B75353
hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000

C:Accession: B75353

R:Wille, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: B75353

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-127 <NR1>

A:Cross-references: GB:AE002021; GB:AE00513; NID:96459573; PIDN:AAE11348.1; PID:9645956

A:Experimental source: strain R1

C:Genetics:

A:Gene: DRI788

A:Map position: 1

C:Superfamily: Deinococcus radiodurans hypothetical protein DRI788

Query Match 45.3%; Score 43; DB 2; Length 127;
Best Local Similarity 69.2%; Pred. No. 29;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 MRVLAVALAVG 13
1 MKVLAVALAVG 13

RESULT 13

S75049

hypothetical protein slr1721 - Synechocystis sp. (strain PCC 6803)

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000

C:Accession: S75049

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

S.

A:Reference number: S74322; MUID:97061201

A:Accession: S75049

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-492 <KAN>

A:Cross-references: EMBL:D90910; GB:AB001339; NID:g1652956; PIDN:BA17911.1; PID:g165295

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 45.3%; Score 43; DB 2; Length 492;
Best Local Similarity 61.1%; Pred. No. 88;
Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 3 VLVLAVALAVG-DQSNL 20
1 VLVLAVALAVG-DQSNL 20
28 VLVLAVALAVG-DQSNL 45

RESULT 14

AB1855

permease protein of iron(III) ABC transporter all0387 [imported] - Anabaena sp. (s

C:Species: Anabaena sp.

A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002

C:Accession: AB1855

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Ir

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tab

DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacteri

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AB1855

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-518 <NR>

A:Cross-references: GB:BA000019; PIDN:BBB72345.1; PID:g17129732; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all0387

Query Match 45.3%; Score 43; DB 2; Length 518;
Best Local Similarity 50.0%; Pred. No. 92;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 2 RVVLAVALAVG-DQSNL 21
1 RVVLAVALAVG-DQSNL 21
423 RVVLAVALAVG-DQSNL 442

RESULT 15

A31770

pheromone-binding protein precursor - tobacco hornworm

C:Species: Manduca sexta (tobacco hornworm)

C>Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 18-Aug-2000

C:Accession: A31770

R:Gyorgyi, T.K.; Roby-Sherkovitz, A.J.; Lerner, M.R.

Proc. Natl. Acad. Sci. U.S.A. 85, 9851-9855, 1988

A:Title: Characterization and cDNA cloning of the pheromone-binding protein from t

A:Reference number: A31770; MUID:89071794

A:Accession: A31770

A:Molecule type: mRNA

A:Residues: 1-168 <GYO>

A:Cross-references: GB:M21797; GB:J04146; NID:g159537; PID:g159538

A:Note: 29-Glu and 120-Val were also found

C:Superfamily: tobacco hornworm pheromone-binding protein

F.1-26/Domain: signal sequence #status predicted <SIG>

F.27-168/Product: pheromone-binding protein #status predicted <MAT>

Query Match 44.2%; Score 42; DB 2; Length 168;
Best Local Similarity 56.2%; Pred. No. 51;
Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 MRVLAVALAVG-D 16
1 MKVLAVALAVG-D 16
8 MKVLAVALAVG-D 23

Search completed: August 6, 2002, 15:03:26
Job time: 156 sec

GenCore version 4..5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 15:01:50 ; Search time 25.03 Seconds
(without alignments)
145.142 Million cell updates/sec

Title: US-09-426-776a-10
Perfect score: 95
Sequence: 1 MRVLVLAVALAVDQSNMG 21

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	89	93.7	1788	13	Q9YXG0 oreochromis
2	52	54.7	1353	13	Q90XA4 cyprinus ca
3	49	51.6	1339	13	Q9M6T2 pinephales
4	46	48.4	1665	13	Q98T86 melanogramm
5	45	47.4	107	10	Q9LEP7 brassica na
6	44	46.3	32	6	Q9BEP6 amblyomma a
7	44	46.3	330	5	Q9BEP6 capra hircu
8	44	45.8	223	6	Q9BEP6 caenorhabdi
9	43.5	45.8	952	5	Q9BEP6 capra hircu
10	43.5	45.8	952	5	Q9BEP6 capra hircu
11	43	45.3	127	16	Q9RTM6 leishmania
12	43	45.3	492	16	Q9RTM6 delnoccoc
13	43	45.3	634	2	Q9S6S8 synchocyst
14	43	45.3	657	2	Q9S6S8 rhizobium 1
15	43	45.3	704	13	Q90865 gallus gall
16	43	45.3	712	10	Q93WP2 chlamydomon

17	42.5	44.7	296	5	Q9V729	Q9V729 drosophila
18	42	44.2	161	5	Q9U4Z5	Q9U4Z5 manduca sex
19	42	44.2	174	17	Q98538	Q98538 methanococc
20	42	44.2	356	2	Q9L3A0	Q9L3A0 sphingomona
21	42	44.2	499	2	Q9RDR5	Q9RDR5 streptomyc
22	42	44.2	751	5	Q9U5K5	Q9U5K5 spodioptera
23	41	43.2	109	41	Q9RWX6	Q9RWX6 delnoccoc
24	41	43.2	247	16	Q9H53	Q9H53 pseudomonas
25	41	43.2	263	5	Q9VPA2	Q9VPA2 drosophila
26	41	43.2	275	10	Q41622	Q41622 tulipa sp.
27	41	43.2	286	5	Q9XZ15	Q9XZ15 drosophila
28	41	43.2	342	16	Q984S6	Q984S6 rhizobium 1
29	41	43.2	481	5	Q21517	Q21517 caenorhabdi
30	41	43.2	821	10	Q9SIT5	Q9SIT5 arabidopsis
31	41	43.2	1274	5	Q9VSO9	Q9VSO9 drosophila
32	41	43.2	1362	13	Q90YN8	Q90YN8 brachydanio
33	40.5	42.6	577	5	Q9VIX7	Q9VIX7 drosophila
34	40	42.1	108	3	Q93917	Q93917 pisolithus
35	40	42.1	111	5	Q62053	Q62053 caenorhabdi
36	40	42.1	208	16	Q9CPJ3	Q9CPJ3 pasteurella
37	40	42.1	233	6	Q9GKR3	Q9GKR3 equus caball
38	40	42.1	264	16	Q9RTG9	Q9RTG9 deinochoc
39	40	42.1	297	8	Q9S0U8	Q9S0U8 hyaloraphid
40	40	42.1	300	16	Q9A6T2	Q9A6T2 caulobacter
41	40	42.1	309	13	Q93605	Q93605 oncorhynch
42	40	42.1	309	13	Q9M6Z4	Q9M6Z4 oncorhynch
43	40	42.1	309	13	Q9PSX4	Q9PSX4 oncorhynch
44	40	42.1	312	16	Q9KQ5	Q9KQ5 bacillus ha
45	40	42.1	325	16	Q98B70	Q98B70 rhizobium 1

ALIGNMENTS

RESULT	ID	Q9YXG0	PRELIMINARY	PRT	1788 AA.
AC	Q9YXG0	Q9YXG0	Q9YXG0		
DT	01-MAY-1999	(TREMBLrel. 10, Created)			
DT	01-MAY-1999	(TREMBLrel. 10, Last sequence update)			
DT	01-JUN-2001	(TREMBLrel. 17, Last annotation update)			
DE	VITELLOGENIN PRECURSOR.				
GN	VIT1.				
OS	Oreochromis aureus (Israeli tilapia).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Neuteleostomi;				
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;				
OC	Acanthomorpha; Acanthopterygii; Perciformes; Labroidae;				
OC	Cichlidae; Oreochromis.				
OX	NCBI_TaxID=47969;				
NP	[1]				
NP	SEQUENCE FROM N.A.				
NP	TISSUE=LIVER:				
RA	Lin E.H., Lam T.J., Ding J.L.;				
RT	"Cloning of Full-Length Oreochromis aureus Vitellogenin cDNA and its				
RT	Deduced Primary Structure."				
RL	Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.				
RM	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Blood:				
RC	MEDLINE=99145366; PubMed=10022768;				
RA	Teo B.Y., Tan N.S., Lin E.H., Lam T.J., Ding J.L.;				
RT	"A novel piscine vitellogenin gene: structural and functional analyses				
RT	of estrogen-inducible promoter."				
RL	Mol. Cell. Endocrinol. 146:103-120(1998).				
DR	EMBL; AF017250; AAD01615.1;				
DR	EMBL; AF072686; AAD48085.1;				
DR	InterPro; IPR001747; Vitellogenin_N.				
DR	InterPro; IPR001846; vwd.				
DR	Pfam; PF01347; Vitellogenin_N; 1.				
DR	Pfam; PF00094; vwd; 1.				
DR	SMART; SM00216; vwd; 1.				
KT	Signal.				
FT	SIGNAL	1	16	POTENTIAL.	

FT CHAIN 17 1788 VITELLOGENIN.
SQ SEQUENCE 1788 AA; 195001 MW; 032A275D45421C9A CRC64;

Query Match 93.7%; Score 89; DB 13; Length 1788;
Best Local Similarity 100.0%; Pred. No. 0.00012;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MRVVLALAVALVAGDSNL 20
DB 1 MRVVLALAVALVAGDSNL 20

RESULT 2

O90XA4 PRELIMINARY; PRT; 1353 AA.

AC O90XA4;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE VITELLOGENIN.
OS Cyprinus carpio (Common carp).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Cyprinus.
OX NCBI_Taxid=7962;
RN [1]
RP SEQUENCE FROM N.A.
RA Chang Y.-S., Lai C., Wu S., Huang F.L.;
RT "Molecular cloning of carp vitellogenin cDNA."
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF414432; AAL07472.1;
SQ SEQUENCE 1353 AA; 148259 MW; 1B26BC0C19DBA76C CRC64;

Query Match 54.7%; Score 52; DB 13; Length 1353;
Best Local Similarity 60.0%; Pred. No. 27;
Matches 12; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 1 MRVVLALAVALVAGDSNL 20
DB 1 MRVVLALAVALVAGDSNL 20

RESULT 3

O9M612 PRELIMINARY; PRT; 1339 AA.

AC O9M612;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE VITELLOGENIN.
GN VTG.
OS Pimphales promelas.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Pimphales.
OX NCBI_Taxid=90988;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=LIVER;
RC Korte J.J., Kahl M.D., Jensen K.M., Pasha M.S., Parks L.G.,
RA LeBlanc G.A., Ankley G.T.;
RT "Fathead Minnow Vitellogenin: Complementary DNA Sequence and Messenger
RT RNA and Protein Expression after 17(beta)-Estradiol Treatment."
RL Environ. Toxicol. Chem. 19:972-981(2000).
DR EMBL; AF130354; AAD23878.1;
DR InterPro: IPR001747; Vitellogenin.N.
DR Pfam: PF01347; Vitellogenin.N.
SQ SEQUENCE 1339 AA; 146274 MW; C5DDF4EC4C3B1A2A CRC64;

Query Match 51.6%; Score 49; DB 13; Length 1339;

Best Local Similarity 60.0%; Pred. No. 75;
Matches 12; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 1 MRVVLALAVALVAGDSNL 20
DB 1 MRVVLALAVALVAGDSNL 20

RESULT 4

O98T86 PRELIMINARY; PRT; 1665 AA.

AC O98T86;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE VITELLOGENIN A.
GN VTGA.
OS Melanogrammus aeglefinus (Haddock).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae;
OC Melanogrammus.
OX NCBI_Taxid=8056;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-2123392; PubMed-11335916;
RA Reith M., Munkholland J., Kelly J., Finn R.N., Flynn H.J.;
RT "lipovitellins derived from two forms of vitellogenin are
RT differentially processed during oocyte maturation in haddock
RT (Melanogrammus aeglefinus)."
RL J. Exp. Zool. 291:58-67(2001).
DR EMBL; AF284035; AAK15158.1;
DR InterPro: IPR001747; Vitellogenin.N.
DR Pfam: PF01347; Vitellogenin.N.
DR Pfam: PF00094; vwd. 1.
DR SMART; SM00216; vwd. 1.
SQ SEQUENCE 1665 AA; 182390 MW; B3433E42B4A19051 CRC64;

Query Match 48.4%; Score 46; DB 13; Length 1665;
Best Local Similarity 52.6%; Pred. No. 2,6e+02;
Matches 10; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 MRVVLALAVALVAGDSN 19
DB 1 MRVVLALAVAGDSN 19

RESULT 5

O9LEP7 PRELIMINARY; PRT; 107 AA.

AC O9LEP7;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PUTATIVE GLYCINE-RICH PROTEIN.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_Taxid=3708;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. N-O-9;
RC Bowers N.L., Trick M.;
RT "Microsinteny at the FCA region between Arabidopsis thaliana and
RT Brassica napus."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ293726; CAC01931.1;
DR InterPro: IPR001525; C5_DNA_meth.
DR PROSITE; PS00095; C5_MTASE.2; UNKNOWN.1.
SQ SEQUENCE 107 AA; 9714 MW; 0A7BE4A01B09B1B3 CRC64;

Matches 11; Conservative 4; Mismatches 5; Indels 1; Gaps 1;
 QY 1 MRVLVLAVALAVAG--DOSNL 20
 Db 1 MRVLVLAVALAVAG--DOSNL 21

RESULT 10
 097198 PRELIMINARY; PRT; 952 AA.
 AC 097198;
 DT 01-MAY-1999 (TREMblrel. 10, Created)
 DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE PROBABLE PROTON MOTIVE ATPASE.
 GN 12385.06
 OS Leishmania major.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 OX NCBI_TaxID=5664;
 RN [1]
 RC STRAIN-FRIEDLIN;
 RA Ivens A.C., Murphy L., Quail M., Harris D., Oliver K., Lawson D.,
 Rajandream M.A., Barrell B.G.,
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC STRAIN-FRIEDLIN;
 RX MEDLINE-9816435; PubMed-9477341;
 RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
 Smith D.F.,
 RT "A physical map of the Leishmania major Friedlin genome."
 RL Genome Res. 8:135-145(1998).
 DR EMBL: AL139794; CAC22655.1;
 DR InterPro: IPR001757; E1-E2_ATPase.
 DR InterPro: IPR000695; HATPase.
 DR InterPro: IPR001454; Hydrolyase.
 DR Pfam: PF00122; E1-E2_ATPase; 1.
 DR Pfam: PF00122; Hydrolyase; 1.
 DR PRINTS: PR00119; CATATPASE.
 DR PRINTS: PR00120; HATPASE.
 SQ SEQUENCE 952 AA; 101677 MW; F81F8046EBB3A1BB CRC64;

Query Match 45.8%; Score 43.5; DB 5; Length 952;
 Best Local Similarity 54.5%; Pred. No. 3.5e+02;
 Matches 12; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

QY 1 MRVLVLAVAL--AVGDOSN 19
 Db 670 LSVLAVALAVAVPRCVCSSN 691
 RESULT 11
 09RTH6 PRELIMINARY; PRT; 127 AA.
 AC 09RTH6;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
 DE HYPOHETICAL 13.8 KDA PROTEIN.
 GN DRI788.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-R1;
 RX MEDLINE-20036896; PubMed-10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,

RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.,
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 RT radiodurans R1."
 RL Science 286:1571-1577(1999).
 DR EMBL: AE002020; AAF1348.1;
 DR TIGR: DRI788;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 127 AA; 13776 MW; F34332516B43A14 CRC64;

Query Match 45.3%; Score 43; DB 16; Length 127;
 Best Local Similarity 69.2%; Pred. No. 57;
 Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MRVLVLAVALA 13
 Db 1 MRVLVLAVALA 13

RESULT 12
 ID P73852 PRELIMINARY; PRT; 492 AA.
 AC P73852;
 DT 01-FEB-1997 (TREMblrel. 02, Created)
 DT 01-FEB-1997 (TREMblrel. 02, Last sequence update)
 DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
 DE HYPOHETICAL 54.4 KDA PROTEIN.
 GN SRI1721.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX MEDLINE-97061201; PubMed-8905231;
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 Miyajima N., Hirosewa M., Sugita M., Sasamoto S., Kimura T.,
 Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
 RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 Tabata S.,
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions."
 RL DNA Res. 3:109-136(1996).
 DR EMBL: D90910; BAA17911.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 492 AA; 54410 MW; 5E216D617270DC6B CRC64;

Query Match 45.3%; Score 43; DB 16; Length 492;
 Best Local Similarity 61.1%; Pred. No. 2.2e+02;
 Matches 11; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 3 VLVLAVLAVALAVGDSNL 20
 Db 28 VLVLAVLAVALAVGDSNL 45
 RESULT 13
 09S6S8 PRELIMINARY; PRT; 634 AA.
 AC 09S6S8;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE FHOB.
 GN FHOB.
 OS Rhizobium leguminosarum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_TaxID=384;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-8401(PRLJ1);
 RX MEDLINE-99231845; PubMed-10217493;
 RA Stevens J.B., Carter R.A., Hussain H., Carson K.C., Dilworth M.J.,
 RA Johnston A.W.B.;
 RT "The fnu genes of Rhizobium leguminosarum, specifying siderophore
 uptake proteins: fnuDCB are adjacent to a pseudogene version of
 fnuA.";
 RT Microbiology 145:593-601(1999).
 RL EMBL: AJ007906; CA07726.1; -;
 DR InterPro: IPR000522; FeCD.
 DR InterPro: IPR002040; Tachykinin.
 DR Pfam: PF01032; FeCD family; 2.
 DR ProDom: PD001557; FeCD; 2.
 DR PROSITE: PS00267; TACHYKININ; UNKNOWN.1.
 SQ SEQUENCE 634 AA; 65246 MW; F9B6EA92FF5DEC1D CRC64;

Query Match 45.3%; Score 43; DB 2; Length 634;
 Best Local Similarity 52.4%; Pred. No. 2.8e+02;
 Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 MRVYLALAVALAVGDSNLG 21
 :|:||||||| |:
 DB 205 LRFIVLALVALAFAVTSAVG 225

RESULT 14
 086463 PRELIMINARY; PRT; 657 AA.
 AC 086463;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE FHUB.
 GN FHUB.
 OS Rhizobium leguminosarum.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OX NCBI_Taxid=384;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-8401(PRLJ1);
 RX MEDLINE-99231845; PubMed-10217493;
 RA Stevens J.B., Carter R.A., Hussain H., Carson K.C., Dilworth M.J.,
 RA Johnston A.W.B.;
 RT "The fnu genes of Rhizobium leguminosarum, specifying siderophore
 uptake proteins: fnuDCB are adjacent to a pseudogene version of
 fnuA.";
 RT Microbiology 145:593-601(1999).
 RL EMBL: AJ007906; CA07725.1; -;
 DR InterPro: IPR000522; FeCD.
 DR InterPro: IPR002040; Tachykinin.
 DR Pfam: PF01032; FeCD family; 2.
 DR ProDom: PD001557; FeCD; 2.
 DR PROSITE: PS00267; TACHYKININ; UNKNOWN.1.
 SQ SEQUENCE 657 AA; 67583 MW; 05D0CB4CC8F39745 CRC64;

Query Match 45.3%; Score 43; DB 2; Length 657;
 Best Local Similarity 52.4%; Pred. No. 2.9e+02;
 Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 MRVYLALAVALAVGDSNLG 21
 :|:||||||| |:
 DB 228 LRFIVLALVALAFAVTSAVG 248

RESULT 15
 090865 PRELIMINARY; PRT; 704 AA.
 AC 090865;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HEPATOCYTE GROWTH FACTOR-LIKE/MACROPHAGE STIMULATING PROTEIN.
 GN HGF/MSP.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_Taxid=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE-96029010; PubMed-7554499;
 RA Thery C., Sharpe M.J., Batley S.J., Stern C.D., Gherardi E.;
 RT "Expression of HGF/SF, HGF/MSP and c-met suggests new functions
 during early chick development.";
 RT dev. Genet. 17:90-101(1995).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC EMBL: X84043; CA58862.1; -;
 DR HSSP: P00747; ICEA.
 DR MEROPS: S01.977; -;
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000001; Kringle.
 DR InterPro: IPR003014; PAN.
 DR InterPro: IPR003609; Pan.app.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00051; Kringle; 4.
 DR Pfam: PF00024; PAN; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PRINTS: PR00018; KRINGLE.
 DR SMART: SM00130; KR; 4.
 DR SMART: SM00473; PAN_AP; 1.
 DR SMART: SM00020; TYP_SPC; 1.
 DR PROSITE: PS00021; KRINGLE_1; 4.
 DR PROSITE: PS00070; KRINGLE_2; 4.
 DR PROSITE: PS50240; TRYPSIN_DOM; 1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 704 AA; 79341 MW; CAB0D8CC41367C37 CRC64;

Query Match 45.3%; Score 43; DB 13; Length 704;
 Best Local Similarity 58.8%; Pred. No. 3.1e+02;
 Matches 10; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 4 LVTLALAVALAVGDSNL 20
 |:|:||||| |:
 DB 7 LLLSLAVALSAGHRSP 23

Search completed: August 6, 2002, 15:03:59
 Job time: 129 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2002, 23:50:44 ; Search time 1894.05 Seconds

(without alignments)
883.885 Million cell updates/sec

Title: US-09-426-776a-11

Perfect score: 80

Sequence: 1 attcaccatccaccagccatg.....gggaccatccacttgggg 80

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_com:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sy:*
12: gb_un:*
13: gb_vl:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_com:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htgo_inv:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	74	92.5	5511	5	AF017250	AF017250 Oreochrom
2	55.8	69.8	11453	5	AF072686	AF072686 Oreochrom
3	43.4	54.2	5166	5	FN070826	FN070826 Fundulus he
4	31.6	39.5	1594	10	AY039762	AY039762 Mus muscu
5	31.6	39.5	1904	10	BC019797	BC019797 Mus muscu
6	30	37.5	137316	2	AC098951	AC098951 Rattus no
7	29.2	36.5	4230	5	AF004678	AF004678 Dario rer
8	29	36.2	42805	9	AC004221	AC004221 Homo sapi
9	28.8	36.0	945	9	HS080197	HS080197 Homo sapi
10	28.8	36.0	4214	9	AK023132	AK023132 Homo sapi
11	28.8	36.0	4813	9	AK024508	AK024508 Homo sapi
12	28.8	36.0	139505	9	HS0591C20	HS0591C20 Homo sapi
13	28.6	35.8	86748	9	AC068717	AC068717 Homo sapi
14	28.4	35.5	10807	6	AR070469	AR070469 Sequence
15	28.4	35.5	10807	6	I47708	I47708 Sequence
16	28.2	35.2	840	3	S66686	S66686 n-syb-vamp
17	28.2	35.2	1113	3	AT069045	AT069045 Drosophi
18	28.2	35.2	50701	2	AC018231	AC018231 Drosophi
19	28.2	35.2	146271	1	SYCSLRB	SYCSLRB Drosophi
20	28.2	35.2	166857	3	AC093192	AC093192 Drosophi
21	28.2	35.2	172173	3	AC010038	AC010038 Drosophi
22	28.2	35.2	257867	3	AC005557	AC005557 Drosophi
23	28.2	35.2	177581	3	AC005472	AC005472 Drosophi
24	28.2	35.2	303191	3	AC099177	AC099177 Rattus no
25	27.8	34.8	113332	2	AC099177	AC099177 Rattus no
26	27.6	34.5	272	4	OALGL1	OALGL1 Ovis arie
27	27.6	34.5	785	4	OALGLR	OALGLR Ovis arie
28	27.6	34.5	824	6	AR002658	AR002658 Sequence
29	27.6	34.5	824	6	AR118412	AR118412 Sequence
30	27.6	34.5	824	6	AR118412	AR118412 Sequence
31	27.6	34.5	1769	4	SHPBLGAL	SHPBLGAL Ovis arie
32	27.6	34.5	7379	4	OALGB	OALGB Ovis arie
33	27.6	34.5	62838	2	AC015633	AC015633 Homo sapi
34	27.4	34.2	5189	5	AP284035	AP284035 Melanogr
35	27.4	34.2	110000	2	AP004361_2	AP004361_2 Continuation (3 of
36	27.4	34.2	139014	8	AP004233	AP004233 Oryza sat
37	27.4	34.2	160030	2	AC106607	AC106607 Rattus no
38	27.4	34.2	161855	2	AC094128	AC094128 Rattus no
39	27.4	34.2	170245	9	HS109F14	HS109F14 Human DNA
40	27.2	34.0	155289	2	OSJN00071	OSJN00071 Oryza sat
41	27.2	34.0	190289	14	MC060315	MC060315 Moluscum
42	27.2	34.0	208408	9	AL450163	AL450163 Human DNA
43	27	33.8	143463	9	AL359312	AL359312 Human DNA
44	27	33.8	152211	4	SSC251829	SSC251829 Sus scro
45	27	33.8	157571	2	AC068255	AC068255 Homo sapi

ALIGNMENTS

RESULT 1
LOCUS AF017250 5511 bp mRNA linear VRT 05-JAN-1999
DEFINITION Oreochromis aureus vitellogenin precursor (Vtgl) mRNA, complete
cgs.
ACCESSION AF017250
VERSION AF017250.1 GI:4102880
KEYWORDS
SOURCE
ORGANISM Oreochromis aureus.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei;
Acanthomorpha; Acanthopterygii; Percormorpha; Perciformes;
Labroidae; Cichlidae; Oreochromis.
1 (bases 1 to 5511)
REFERENCE
AUTHORS Lim,E.H., Lam,T.J. and Ding,J.L.
TITLE Cloning of Full-length Oreochromis aureus Vitellogenin cDNA and its
Deduced Primary Structure
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 5511)
AUTHORS Lim,E.H., Lam,T.J. and Ding,J.L.
TITLE Direct Submission
JOURNAL Submitted (05-AUG-1997) School of Biological Sciences, National

misc_feature	58..117	/note="encodes region of N-terminal amino acids of a 69 kDa-yolk protein (YP 69) that were experimentally determined by Edman degradation"
mat_peptide	58..5073	
misc_feature	3235..3660	/product="vitellogenin II"
		/note="encodes Yrg polyserine domain found in phosphoserine rich yolk proteins termed phosvitin and/or phosvettes"
polya_signal	5135..5140	
BASE COUNT	1360 a 1358 c 1228 g 1220 t	
ORIGIN		
Query Match	54.2%;	Score 43.4; DB 5; Length 5166;
Best Local Similarity	76.8%;	Pred. No. 0.00021;
Matches	53; Conservative	0; Mismatches 16; Indels 0; Gaps 0;
QY	7	atccccagccatgggggtgcttgactacgtcttgctggtgcttcgcagtggggacc 66
Db	2	ATTACACGACCCWAGGAGGCTTGTCGTGGCTCTCCTACGTGTGGCCCTTGTCGGGGAACC 61
QY	67	agtcacat 75
Db	62	AGGTGAGCT 70
RESULT	4	
AY039762	1594 bp	mRNA linear ROD 11-OCT-2001
LOCUS	AY039762	Mus musculus 3-beta-hydroxysterol delta-24 reductase mRNA, complete
DEFINITION	AY039762.1	GI:16037746
ACCESSION	AY039762	
VERSION	AY039762.1	
KEYWORDS		
SOURCE		house mouse.
ORGANISM		Mus musculus
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS		1 (bases 1 to 1594)
TITLE		Waterham,H.R., Koster,J., Romeijn,G.J., Hennekam,R.C., Vreken,P., Andersson,H.C., FitzPatrick,D.R., Kelley,R.I. and Wanders,R.J.
		Mutations in the 3beta-Hydroxysterol Delta24-Reductase Gene Cause Desmosterolosis, an Autosomal Recessive Disorder of Cholesterol Biosynthesis
JOURNAL		Am. J. Hum. Genet. 69 (4), 685-694 (2001)
MEDLINE		21426336
PUBMED		11519011
REFERENCE		2 (bases 1 to 1594)
AUTHORS		Waterham,H.R. and Koster,J.
TITLE		Direct Submission
JOURNAL		Submitted (13-JUN-2001) Lab. Genet. Metab. Dis. (P0-224), Academic Medical Center, University of Amsterdam, Meibergdreef 9, Amsterdam 1105 AZ, The Netherlands
FEATURES		Location/Qualifiers
source		1..1594
		/organism="Mus musculus"
		/db_xref="taxon:10090"
CDS		18..1574
		/function="catalyzes the reduction of the C24-C25 double bond of desmosterol, or other delta 24 sterol intermediates, to produce cholesterol in the cholesterol biosynthetic pathway"
		/note="sterol reductase; 24-dehydrocholesterol reductase; dhcr24"
		/codon_start=1
		/product="3-beta-hydroxysterol delta-24 reductase"
		/protein_id="AAK2106.1"
		/db_xref="GI:16037746"

BASE COUNT	334 a	455 c	471 g	334 t
ORIGIN				
Query Match	Best Local Similarity	39.5%;	Score 31.6; DB 10;	Length 1594;
Matches	40; Conservative	0; Mismatch	2.3; Indels	0; Gaps 0;
Qy	2	ttcacatccaccagcgcattgagtgtctgtactacgtcttgcgtgtgctcgcc 55 Db 94 TTCATCATCCACAGCGCTGGGCTTCGTCCGTCTTCTTGCTGCCGCTCACC 147		
RESULT 5	BC019797	1904 bp	mRNA	linear ROD 22-JAN-2002
LOCUS	Mus musculus, 24-dehydrocholesterol reductase, clone MGC:30547			
DEFINITION	IMAGE:5054108, mRNA, complete cds.			
ACCESSION	BC019797			
VERSION	BC019797.1	GI:18044183		
KEYWORDS	MGC.			
SOURCE	house mouse.			
ORGANISM	Mus musculus. Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 1904)			
REFERENCE	Strausberg, R. Direct Submission Submitted (19-DEC-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA			
AUTHORS	NIH-MGC Project URL: http://mgc.nci.nih.gov			
TITLE	Contact: MGC help desk			
JOURNAL	Email: cgabs-remail.nih.gov Tissue Procurement: Jeffrey E. Green, M.D. CDNA Library Preparation: Life Technologies, Inc. cDNA library Arrayed by: The I.M.A.G.E. Consortium (ILNI) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: amgebcm.tmc.edu Gunaratne, P.H.; Garcia, A.M.; Lu, X.; Huijck, S.W.; Hale, S.M.; Yoon, V.S.; Kowis, C.R.; Lawrence, S.; Martin, R.G.; Muzny, D.M.; Richards, S.; Gibbs, R.A.			
REMARK	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov			
COMMENT	Series: IRAP Plate: 41 Row: 1 Column: 8 This clone was selected for full length sequencing because it passed the following selection criteria: GenomScan gene prediction.			
FEATURES	Location/Qualifiers			
source	1..1904			
	/organism="Mus musculus"			
	/db_xref="taxon:10090"			
	/clone="MGC:30547 IMAGE:5054108"			
	/tissue_type="Liver, normal. 5 month old male mouse."			
	/clone_id="NCI_CGAP_L19"			
	/lab_host="DH10B"			
	/note="Vector: pCMV-SPORT6"			

CDS

74..1624
/codon_start=1
/product="24-dehydrocholesterol reductase"
/protein_id="AAH19797.1"
/db_xref="GI:18044184"
/translation="MEPAVSLVAVCALFLVAVKGLFEVLIHQRVFCVCLLPST
IFDIYVYRANVAVKLSAPRLHQRVADIOKQVREMEQSCFPMCGRMVLVSL
RVGKRTKKNMINIMNDILEVDTRKQIVRVEPLVMSGCVTALNSIGMTPLVPELD
DLVYGLMGCTGSISSSHKGLFQHCYAVELIADSPFRCTPSENSDILFVAVMSC
GTGFLVAAERITIPAKRYVLRREPVGALICRCKTRESORLENHVEGLISLDE
AVMTGVTDDVPEPSKLSISYTPMFKFVANTLKTNRGLEYITPLRHYHRTS
IFMELKODIIPGNMIFRYLFGWVPPKISLKLQGTLEKLEQHHVDDMLVPMK
CMQALHTFQNDIHYPIWLCPELIPISQPGVHPGDAELEYVDGAYGEPRVKEEA
RSCMRKLEKFRVSHGFQMLVADCYMNRREEMEDGSLYHKLKRLQCCQAFPEYVD
KICKARRH"

BASE COUNT 403 a 553 c 562 g 386 t

ORIGIN

Query Match 39.5%; Score 31.6; DB 10; Length 1904;
Best Local Similarity 74.1%; Pred. No. 2.3;
Matches 40; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Ox 2 ttacatccaccagccatgaggtgcttactagcttctgtgctctgc 55
|||||
Db 150 ttctatccacacagcgcgtgctgtctgtcttcttctgctgcgc 203

RESULT 6
AC098951 137316 bp DNA linear HTG 20-DEC-2001
LOCUS Rattus norvegicus clone CH230-173E15, *** SEQUENCING IN PROGRESS
DEFINITION *** 68 unordered pieces.
AC098951.2 GI:17973735
VERSION HTG; HTGS_PHASE1.
KEYWORDS Norway rat.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 137316)
Muzny D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C.,
Alsbrooks S.L., Amaralunga H.C., Are J.R., Banks T., Barbarta J.,
Benton J., Blum K., Blankenburg K., Bonnin D., Bouck J.,
Boyte S., Brieva M., Brown E., Brown M., Bryant N.P., Buhay C.,
Burke P., Burkett C., Burrell K.L., Byrd N.C., Carron T.F.,
Cartier M., Cavazos S.R., Chacko J., Chavez D., Chen G., Chen R.,
Chen Z., Chowdhry I., Christopoulos C., Cleveland C.D., Cox C.,
Coyle M.D., Dathorne S.R., David R., Davila M.L., Davis C.,
Day-Carroll L., Dederich D.A., Delaney K.R., Delgado O.,
Dem A.L., Ding Y., Dinh H.H., Douthwaite K.J., Diaper H.,
Dujan-Rocha S., Durbin K.J., Earmhart C., Edgar D., Edwards C.C.,
Elhaj C., Escotto M., Falls T., Ferraguto D., Flagg N., Ford J.,
Foster P., Frantz P., Gabisi A., Gao J., Garcia A., Garner T.,
Gara N., Gill R., Gorrell J.H., Guevara W., Gunaratne P., Hale S.,
Hamilton K., Harris C., Harris K., Hart M., Haylak P., Hayes A.,
Hernandez J., Hernandez O., Hodgson A., Hogues M., Holloway C.,
Hollins B., Homsl F., Howard S., Huber J., Huliy S., Hume J.,
Jackson L.E., Jacobson B., Jia Y., Johnson R., Jollivet S.,
Joudah S., Karlsson E., Kelly S., Khan U., King L., Korvah J.,
Kovar C., Kratovic J., Kureshi A., Landry N., Leal B., Lewis L.C.,
Lewis L., Li J., Li Z., Lichtarge O., Lieu C., Liu C., Liu W.,
Loulseged H., Lozano R.J., Lu X., Lucier A., Lucier R., Luna R.,
Ma J., Maheshwari M., Mapua P., Martin R., Martindale A.,
Martinez E., Massey E., Mawhinney E., McLeod M.P., Meador M.,
Mel G., Metzger M., Miner G., Miner Z., Mitchell T., Mohabhat K.,
Morgan M., Morris S., Moser M., Neal D., Newton J., Newson N.,
Nguyen A., Nguyen N., Nguyen N., Nickerson E., Nwokwenkwo S.,
Ogulu M., Okunolu G., Oragunye N., Orvedo R., Pace A., Payton B.,
Peery J., Perez L., Peters J., Pickens R., Primus E., Pu L.L.,
Quiles S., Ren Y., Rives M., Rojas A., Rojupokan I., Rolfe M.,
Ruiz S., Savery G., Scherer S., Scott G., Shen H., Shooshbari N.,

Siison, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Swale, A., Taber, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, Y., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wlezyk, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission
Unpublished
2 (bases 1 to 137316)
Worley, K.C.
Direct Submission
Submitted (07-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Dec 20, 2001 this sequence version replaced gi:16756145.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GIUG
Center clone name: CH230-173E15
----- Summary Statistics
Assembly program: Phrap; version 0.990329First call to
findPhrapList
Consensus quality: 96465 bases at least Q40
Consensus quality: 104260 bases at least Q30
Consensus quality: 109517 bases at least Q20
Estimated insert size: 77121; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-tp estimation
Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1
* 4489: contig of 4488 bp in length
* 4588: gap of unknown length
* 4589: contig of 3880 bp in length
* 4589: gap of unknown length
* 8569: contig of 3368 bp in length
* 11937: gap of unknown length
* 12037: contig of 3091 bp in length
* 15127: gap of unknown length
* 15227: gap of unknown length
* 15228: contig of 3746 bp in length
* 18974: gap of unknown length
* 19073: gap of unknown length
* 22125: contig of 3052 bp in length
* 22126: gap of unknown length
* 22225: gap of unknown length
* 26029: contig of 3804 bp in length
* 26030: gap of unknown length
* 26130: contig of 2976 bp in length
* 29105: gap of unknown length
* 29205: gap of unknown length
* 32439: contig of 3234 bp in length
* 32440: gap of unknown length
* 32540: contig of 3993 bp in length
* 36533: gap of unknown length
* 36632: gap of unknown length
* 39121: contig of 2489 bp in length
* 39221: gap of unknown length
* 39222: contig of 2738 bp in length
* 41960: gap of unknown length
* 42059: contig of 2834 bp in length
* 44893: gap of unknown length
* 44894: gap of unknown length
* 44993: contig of 2244 bp in length
* 44994: gap of unknown length
* 47237: gap of unknown length
* 47238: gap of unknown length


```

repeat_region
  /rpt_family="GC-rich"
  409. .4301
repeat_region
  /rpt_family="GC-rich"
  4454. .4513
repeat_region
  /rpt_family="CGGCGn"
  complement(4576. .4650)
  /rpt_family="GC-rich"
  complement(5872. .6046)
  /note="Predicted exon, program: grail2exons_human_1.3,
  frame: 2, quality: excellent, score: 80.000"
  /note="(5966. .6046)
  complement(5966. .6046)
  /note="DDS similarity to overlapping ESTs:
  (4584. .4662) W52572 zc45h09.s1 Soares senescent
  fibroblasts NbHSF Homo sapiens cDNA clone 325313 3'
  (271. .193); 96% identity.-(4662. .4533) W52571 zc45h09.r1
  Soares senescent fibroblasts NbHSF Homo sapiens cDNA clone
  325313 5' (253. .393); 67% identity.
  "
  /complement(6557. .6585)
  /rpt_family="NR"
  /rpt_family="NR"
  complement(7506. .7663)
  /rpt_family="MERS58B"
  complement(8197. .8407)
  /note="Predicted exon, program: grail2exons_human_1.3,
  frame: 0, quality: good, score: 74.000-DDS similarity to
  overlapping ESTs:
  (7023. .6813) W52571 zc45h09.r1 Soares senescent
  fibroblasts NbHSF Homo sapiens cDNA clone 325313 5'
  (45. .252); 90% identity.-(6991. .6813) W52572 zc45h09.s1
  Soares senescent fibroblasts NbHSF Homo sapiens cDNA
  clone325313 3' (457. .272); 90% identity."
  /complement(8962. .9102)
  /note="Predicted exon, program: grail2exons_human_1.3,
  frame: 1, quality: excellent, score: 87.000-(7621. .7518)
  DDS similarity to W52571 zc45h09.r1 Soares senescent
  fibroblasts NbHSF Homo sapiens cDNA clone 325313 5'
  (1. .44); 98% identity."
  /complement(9342. .9598)
  /note="Predicted exon, program: grail2exons_human_1.3,
  frame: 2, quality: excellent, score: 77.000"
  complement(9772. .9916)
  /note="Predicted exon, program: grail2exons_human_1.3,
  frame: 2, quality: excellent, score: 100.000"
  complement(9941. .10050)
  /rpt_family="GC-rich"
  10172. .10260
  /rpt_family="(CAGCA)n"
  complement(10272. .10336)
  /note="Predicted exon, program: grail2exons_human_1.3,
  frame: 1, quality: excellent, score: 96.000"
  complement(10550. .10685)
  /note="Predicted exon, program: grail2exons_human_1.3,
  frame: 1, quality: excellent, score: 100.000"
  complement(10606. .10659)
  /rpt_family="(TGC)n"
  complement(10767. .10853)
  /note="Predicted exon, program: grail2exons_human_1.3,
  frame: 2, quality: excellent, score: 88.000"
  10813. .10864
  /rpt_family="(TGC)n"
  11545. .11656
  /rpt_family="(CGC)n"
  complement(12493. .12523)
  /rpt_family="GC-rich"

```

```

repeat_region      12553..12627
                    /rpt_family="(CGG)n"
repeat_region      12646..12913
                    /rpt_family="GC-rich"
repeat_region      complement(13948..14249)
                    /rpt_family="Alu/b"
repeat_region      complement(15011..15068)
                    /rpt_family="GC-rich"
repeat_region      complement(15443..15762)

Query Match      36.28; Score 29; DB 9; Length 42805;
Best Local Similarity 71.7%; Pred. No. 22;
Matches 38; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 23 gggtctgttactagctctgtctgtctgcacagtgaggagaccacact 75
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  ||
Db 29726 gctcctgcacacagactgcactgctcctgcctgcctggagacagctcact 29674

RESULT 9
LOCUS      HSM801971                      945 bp    mRNA    linear    PRI 18-FEB-2000
DEFINITION Homo sapiens mRNA: cDNA DKFZp434N1429 (from clone DKFZp434N1429);
ACCESSION   AL137301
VERSION     AL137301.1 GI:6807764
KEYWORDS
SOURCE      human.
            Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 945)
            Koehler, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
            Direct Submission
            Submitted (15-JAN-2000) MIPs, Am Klopfer spitze 18a, D-82152
            Martinsried, GERMANY
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkz-heidelberg.de;
            sequenced by BMF (Biomedical Research Center at the Charité,
            Berlin/Germany) within the cDNA sequencing consortium of the German
            Genome Project.
            This clone (DKFZp434N1429) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
            information about the clone and the sequencing project is available
            at http://www.mips.biochem.mpg.de/proj/cDNA/.
            Location/Qualifiers
                1..945
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="DKFZp434N1429"
                /tissue_type="testis"
                /clone_id="434 (synonym: htes3). Vector pSport1; host
                DH10B; sites NotI + SalI
                /dev_stage="adult"
                1..408
                /gene="DKFZp434N1429"
                <1..408
                /gene="DKFZp434N1429"
                /note="unknown"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="CAB70686.1"
                /db_xref="GI:6807765"
                /translation="LNMNQKLHFTCYQOQFRSSHAIVSGGVSYCYHMCICITIAAG
                VSFILASGEVSTIRKGSSEAKSGLVGGGPGAGFSPRCETGSENEPTVNCGCVPRDL
                PRPPALASALGGVPIIGAVCHAHLTQVCVPV"
                890..895
                polyA_signal
                polyA_site      223 a      216 c      282 g      224 t
BASE COUNT      223 a      216 c      282 g      224 t
ORIGIN

```


Query Match	36.0%	Score 28.8	DB 9	Length 945
Best Local Similarity	60.0%	Pred. No. 23		
Matches 48	Conservative 0	Mismatches 32	Indels 0	Gaps 0
QY 1	attacatccaccgcgatgaggtgctgtctgtacgtctgtctgtgtgtcgtcgcagtgg 60			
Db 702	AATCAAAATGCACACAGCATGGAAGCTGCTTGTTCATGCACAGCTGTTTGTCACTGGCGTCTT 761			
QY 61	ggagaccagttccaaacttgagg 80			
Db 762	GGAATCCGACCGACCTGTGGGG 781			
RESULT 10				
AK023132	4214 bp	linear	PRI 29-SEP-2000	
LOCUS	Homo sapiens CDNA FLJ13070 fls, clone NT2RP3001777.			
DEFINITION	AK023132			
ACCESSION	AK023132.1 GI:10434913			
VERSION	oligo capping: fls (full insert sequence).			
KEYWORDS	Homo sapiens teratocarcinoma cell_line:NT2 CDNA to mRNA,			
SOURCE	clone_11b:NT2RP3 clone:NT2RP3001777.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
TITLE	1 (sites)			
JOURNAL	Isogai,T., Ota,T., Hayashi,K., Sugiyama,T., Otsuki,T., Suzuki,Y.,			
REFERENCE	Isogai,T., Nagai,K., Sugano,S., Shiratori,A., Sudo,H.,			
AUTHORS	Nishikawa,T., Hosoi,T., Kaku,Y., Kodaira,H., Kondo,H.,			
TITLE	Magatsuma,M., Takahashi,M., Chiba,Y., Ishida,S., Murakawa,K.,			
JOURNAL	Omori,Y., Takiguchi,S., Watanabe,S., Kimura,K., Murakami,K.,			
REFERENCE	Ishii,S., Kawai,Y., Saito,K., Yamamoto,J., Wakamatsu,A.,			
AUTHORS	Nakamura,Y., Nagahashi,K., Masuno,Y., Minomiya,K. and Iwayanagi,T.			
TITLE	NEDO human cDNA sequencing project			
JOURNAL	Unpublished (2000)			
REFERENCE	2 (bases 1 to 4214)			
AUTHORS	Isogai,T. and Otsuki,T.			
TITLE	Direct Submission			
JOURNAL	Submitted (23-NOV-2000) to the DDBJ/EMBL/GenBank databases. Takao			
REFERENCE	Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana,			
AUTHORS	Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp,			
TITLE	Tel:81-438-52-3951, Fax:81-438-52-3952)			
JOURNAL	NEDO human cDNA sequencing project supported by Ministry of			
REFERENCE	International Trade and Industry of Japan; cDNA full insert			
AUTHORS	sequencing: Research Association for Biotechnology; cDNA library			
TITLE	construction, 5'- 6' 3'-end one pass sequencing and clone selection			
JOURNAL	Helix Research Institute (supported by Japan Key Technology Center			
REFERENCE	etc.) and Department of Virology, Institute of Medical Science,			
AUTHORS	University of Tokyo.			
TITLE	Location/Qualifiers			
JOURNAL	1..4214			
REFERENCE	/organism="Homo sapiens"			
AUTHORS	/db_xref="taxon:9606"			
TITLE	/cell_line="NT2"			
JOURNAL	/cell_type="teratocarcinoma"			
REFERENCE	/clone_11b="NT2RP3"			
AUTHORS	/note="Cloning vector: pME18SFLJ3-mRNA from NT2 neuronal			
TITLE	precursor cells after 2-weeks retinoic acid (RA)			
JOURNAL	induction."			
BASE COUNT	779 a 1120 c 1278 g 1037 t			
ORIGIN				
Query Match	36.0%	Score 28.8	DB 9	Length 4214
Best Local Similarity	60.0%	Pred. No. 23		
Matches 48	Conservative 0	Mismatches 32	Indels 0	Gaps 0
QY 1	attacatccaccgcgatgaggtgctgtctgtacgtctgtctgtgtgtcgtcgcagtgg 60			
Db 4009	AATCAAAATGCACACAGCATGGAAGCTGCTTGTTCATGCACAGCTGTTTGTCACTGGCGTCTT 4068			
QY 61	ggagaccagttccaaacttgagg 80			

[illegible]

Query Match	Best Local Similarity	Score	DB	Length
Best Local Similarity 60.0%;	Pred. No. 27;			139505;
Matches 48; Conservative 0;	Mismatches 32;	Indels 0;	Gaps 0;	
QY 1 attaccatccacagcatgaggtgctgtactagctctgtgtgagctcgacagtgg 60				
Db 76746 AATCAATAGCAGCAGCATGAGAGCTGCTGTCATGACATGCTTTGTCTACTGGCGTCT 76805				
QY 61 gggaccagtcacacttgagg 80				
Db 76806 GGAATCGAGCCGCTTGGGG 76825				
RESULT 13	AC068717/c	86748 bp	DNA	linear
LOCUS	AC068717	86748 bp	DNA	linear
DEFINITION	Homo sapiens BAC clone RP11-13K9 from 4, complete sequence.			
ACCESSION	AC068717.5	GI:14589657		

KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
TITLE	1 (bases 1 to 86748)
JOURNAL	Sulston,J.E. and Waterston,R.
MEDLINE	Toward a complete human genome sequence
REFERENCE	Genome Res. 8 (11), 1097-1108 (1998)
AUTHORS	99063792
TITLE	2 (bases 1 to 86748)
JOURNAL	Radienko,M. and Abbot,J.A.
REFERENCE	The sequence of Homo sapiens BAC clone RP11-13X9
AUTHORS	Unpublished
TITLE	3 (bases 1 to 86748)
JOURNAL	Waterston,R.H.
REFERENCE	Direct Submission
AUTHORS	Submitted (07-MAY-2000) Genome Sequencing Center, Washington
TITLE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,
JOURNAL	MO 63108, USA
REFERENCE	4 (bases 1 to 86748)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (03-JUL-2001) Genome Sequencing Center, Washington
REFERENCE	University School of Medicine, 4444 Forest Park Parkway, St. Louis,
AUTHORS	MO 63108, USA
TITLE	5 (bases 1 to 86748)
JOURNAL	Waterston,R.
REFERENCE	Direct Submission
AUTHORS	Submitted (07-NOV-2001) Department of Genetics, Washington
TITLE	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL	On Jul 3, 2001 this sequence version replaced gi:13794249.
COMMENT	

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Oosegawa, K., Moon, P. Y., Zhao, B., Frengen, E., Tateno, M., Cataneese, J. J., and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://Dpacpc.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-568024, 2000 bp overlap. The clone sequenced to the right is RP11-818H4. Actual start of

TITLE Protein C production in non-human transgenic mammals
 JOURNAL Patent: US 5905185-A 5 18-MAY-1999;
 FEATURES Location/Qualifiers
 source 1..10807
 BASE COUNT 2214 a 3066 c 2985 g 2542 t
 ORIGIN

Query Match 35.5%; Score 28.4; DB 6; Length 10807;
 Best Local Similarity 62.9%; Pred. No. 33;
 Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 9 ccaccagcatgaggtgctgtactagctctgtgctcgcagtgaggagaccag 68
 DB 4249 CCTGCAGCCATGAGTGCCTCCTGCTTGCCCTGGCCCTGCCTGCTGTCAG 4308

QY 69 tccaactgg 78
 DB 4309 GCCATCATCG 4318

RESULT 15
 LOCUS 147708 147708 10807 bp DNA linear PAT 07-OCT-1997
 DEFINITION Sequence 7 from patent US 5639940.
 ACCESSION 147708
 VERSION 147708.1 GI:2471673
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 10807)
 AUTHORS Garner, I., Dalrymple, M.L., Prunkard, D.E. and Foster, D.C.
 TITLE Production of fibrinogen in transgenic animals
 JOURNAL Patent: US 5639940-A 7 17-JUN-1997;
 FEATURES Location/Qualifiers
 source 1..10807
 BASE COUNT 2214 a 3066 c 2985 g 2542 t
 ORIGIN

Query Match 35.5%; Score 28.4; DB 6; Length 10807;
 Best Local Similarity 62.9%; Pred. No. 33;
 Matches 44; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 9 ccaccagcatgaggtgctgtactagctctgtgctcgcagtgaggagaccag 68
 DB 4249 CCTGCAGCCATGAGTGCCTCCTGCTTGCCCTGGCCCTGCCTGCTGTCAG 4308

QY 69 tccaactgg 78
 DB 4309 GCCATCATCG 4318

Search completed: August 7, 2002, 01:04:15
 Job time: 4411 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 14:59:55 ; Search time 13.04 Seconds
(without alignments)
39.336 Million cell updates/sec

Title: US-09-426-776A-10

Perfect score: 95

Sequence: 1 MRYVYLAVALAVALGDSNIG 21

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:.*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:.*
3: /cgn2_6/prodata/2/1aa/5A_COMB.pep:.*
4: /cgn2_6/prodata/2/1aa/5B_COMB.pep:.*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:.*
6: /cgn2_6/prodata/2/1aa/Backfiles1.pep:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	46.3	100	1	US-08-384-367-2
2	44	46.3	629	1	US-08-250-740-33
3	44	46.3	629	1	US-07-695-4728-2
4	43.5	45.8	222	2	US-08-391-743A-2
5	42	44.2	683	1	US-07-878-960-2
6	42	44.2	683	2	US-08-477-396A-17
7	40	42.1	361	3	US-09-120-365-75
8	40	42.1	361	4	US-09-515-039-75
9	40	42.1	507	1	US-08-484-493-12
10	40	42.1	507	2	US-08-484-494-12
11	40	42.1	507	2	US-08-345-212-12
12	40	42.1	507	4	US-09-249-003-12
13	40	42.1	509	1	US-08-445-586-8
14	39	41.1	777	2	US-08-477-396A-4
15	38	40.0	374	1	US-08-095-726-14
16	38	40.0	374	1	US-08-096-623A-14
17	38	40.0	374	4	US-08-821-994-70
18	37	38.9	184	4	US-08-969-317-2
19	37	38.9	214	1	US-07-953-230A-11
20	37	38.9	317	4	US-08-949-155-6
21	37	38.9	748	2	US-08-920-234-2
22	37	38.9	748	2	US-08-937-931-4
23	37	38.9	748	2	US-08-285-502-4
24	37	38.9	748	4	US-09-030-335-9
25	37	38.9	748	4	US-09-709-126-4
26	37	38.9	799	4	US-09-030-335-4
27	37	38.9	1241	4	US-09-040-774-2

28	37	38.9	1276	4	US-08-937-236-3	Sequence 3, Appl1
29	37	38.9	1277	4	US-08-937-236-6	Sequence 6, Appl1
30	37	38.9	1291	4	US-08-569-214-3	Sequence 3, Appl1
31	37	38.9	1291	4	US-08-937-236-2	Sequence 2, Appl1
32	37	38.9	1292	4	US-08-569-214-5	Sequence 5, Appl1
33	37	38.9	1292	4	US-08-569-214-6	Sequence 6, Appl1
34	37	38.9	1292	4	US-08-937-236-5	Sequence 5, Appl1
35	37	38.9	1295	4	US-08-569-214-2	Sequence 2, Appl1
36	36	37.9	96	1	US-08-518-878B-35	Sequence 35, Appl1
37	36	37.9	96	1	US-08-294-522B-34	Sequence 34, Appl1
38	36	37.9	96	2	US-08-807-861A-35	Sequence 35, Appl1
39	36	37.9	96	2	US-08-470-868A-35	Sequence 35, Appl1
40	36	37.9	96	2	US-09-210-681-35	Sequence 35, Appl1
41	36	37.9	96	3	US-08-946-719A-35	Sequence 35, Appl1
42	36	37.9	148	2	US-08-888-497-36	Sequence 36, Appl1
43	36	37.9	148	4	US-09-362-230-36	Sequence 36, Appl1
44	36	37.9	148	5	PCT-US94-0726-36	Sequence 36, Appl1
45	36	37.9	197	4	US-08-996-408-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1

US-08-384-367-2
; Sequence 2, Application US/08384367

Patent No. 5629469

GENERAL INFORMATION:
APPLICANT: Deluca-Flaherty, Camille

APPLICANT: Chan, Victor J.
APPLICANT: Scarafia C., Lilliana E.

APPLICANT: Brunke, Karen J.
TITLE OF INVENTION: NOVEL THIOL PROTEASE INHIBITOR

NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sandoz Agro, Inc.

STREET: 975 California Avenue
CITY: Palo Alto

STATE: CA
COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,367

FILING DATE: 10-FEB-1995
CLASSIFICATION: 530

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/208,571

FILING DATE: 10-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Marcus-Wyner, Lynn

REGISTRATION NUMBER: 34,869
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/354-3588

TELEFAX: 415/857-1125
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 100 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-384-367-2

Query Match

Best Local Similarity 46.3%; Score 44; DB 1; Length 100;

Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 RVLVLAVALAVGDOS 18
: | | | | | :
Db 4 KVEVLALIVAVASQT 20

RESULT 2
US-08-250-740-33

; Sequence 33, Application US/08250740
; Patent No. 5686240
; GENERAL INFORMATION:
; APPLICANT: Schuchman, Edward H.
; APPLICANT: Desnick, Robert J.
; TITLE OF INVENTION: Acid Sphingomyelinase Gene and Diagnosis
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/250,740
; FILING DATE: 27-MAY-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 6923-038
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-08-250-740-33

Query Match 46.3%; Score 44; DB 1; Length 629;
Best Local Similarity 71.4%; Pred. No. 35;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 VLVLAVALAVGD 16
: | | | | | :
Db 36 VLAVALALALALSD 49

RESULT 3
US-07-695-472B-2

; Sequence 2, Application US/07695472B
; Patent No. 5773278
; GENERAL INFORMATION:
; APPLICANT: Schuchman, Edward H.
; APPLICANT: Desnick, Robert J.
; TITLE OF INVENTION: The Acid Sphingomyelinase Gene and
; TITLE OF INVENTION: Diagnosis of Niemann-Pick Disease
; NUMBER OF SEQUENCES: 36
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/695,472B
; FILING DATE: 19910503
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6923-014
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 790864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-07-695-472B-2

Query Match 46.3%; Score 44; DB 1; Length 629;
Best Local Similarity 71.4%; Pred. No. 35;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 VLVLAVALAVGD 16
: | | | | | :
Db 36 VLAVALALALALSD 49

RESULT 4
US-08-391-743A-2

; Sequence 2, Application US/08391743A
; Patent No. 5843705
; GENERAL INFORMATION:
; APPLICANT: DiTullio, Paul A.; Meade, Harry; Cole, Edward S.
; TITLE OF INVENTION: TRANSGENETICALLY PRODUCED ANTITHROMBIN III
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/391,743A
; FILING DATE: 21-FEB-1995
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-391-743A-2

Query Match 45.8%; Score 43.5; DB 2; Length 222;
Best Local Similarity 52.4%; Pred. No. 13;
Matches 11; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 1 MRVTLALVALAVG-DOSSL 20
1:|||||:||||:|1|
DB 1 MKVLIACVLAIAAREEL 21

RESULT 5
US-07-878-960-2
Sequence 2, Application US/07878960

GENERAL INFORMATION:
APPLICANT: Purchio, Anthony F.
APPLICANT: Skonier, John
APPLICANT: Neubauer, Michael G.
TITLE OF INVENTION: TGF-BETA INDUCED GENE AND PROTEIN
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bristol-Myers Squibb Company
STREET: 3005 First Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98121

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/878, 960
FILING DATE: 05-MAY-1992
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/833,835
FILING DATE: 05-FEB-1992

ATTORNEY/AGENT INFORMATION:
NAME: 530
NAME: Sorrentino, Joseph M.
REGISTRATION NUMBER: 32,598
REFERENCE/DOCKET NUMBER: ON0092-
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206/728-4800
TELEFAX: 206/727-3601

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 683 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: LUNG
CELL TYPE: ADENOCARCINOMA
CELL LINE: A549
US-07-878-960-2

Query Match 44.2%; Score 42; DB 1; Length 683;
Best Local Similarity 55.6%; Pred. No. 79;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 VLVIALVALAVGDSNL 20

Db 5 VRLIALALALGPATL 22
1:|||||:||||:|1|

RESULT 6
US-08-477-396A-17
Sequence 17, Application US/08477396A
Patent No. 5872235

GENERAL INFORMATION:
APPLICANT: Chen, Ian Bo
APPLICANT: Bao, Shideng
APPLICANT: Liu, Yuan
TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,396A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/146,488
FILING DATE: 29-OCT-1993
APPLICATION NUMBER: US 08/448,388
FILING DATE: 28-MAY-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/12502
FILING DATE: 31-OCT-1994

ATTORNEY/AGENT INFORMATION:
NAME: Helme, Holliday C.
REGISTRATION NUMBER: 34,346
REFERENCE/DOCKET NUMBER: DECI-333BX
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-2290
TELEFAX: (617) 451-0313
INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:
LENGTH: 683 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
US-08-477-396A-17

Query Match 44.2%; Score 42; DB 2; Length 683;
Best Local Similarity 55.6%; Pred. No. 79;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 VLVIALVALAVGDSNL 20
1:|||||:||||:|1|
DB 5 VRLIALALALGPATL 22

RESULT 7
US-09-120-365-75
Sequence 75, Application US/09120365
Patent No. 6103514
GENERAL INFORMATION:

APPLICANT: Natori, Shunji
TITLE OF INVENTION: NEW PROTEASE
FILE REFERENCE: 32290-144749
CURRENT APPLICATION NUMBER: US/09/120,365
CURRENT FILING DATE: 1998-07-22
EARLIER APPLICATION NUMBER: JP 9-333 474
EARLIER FILING DATE: 1997-11-18
NUMBER OF SEQ ID NOS: 101
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 75
LENGTH: 361
TYPE: PRT
ORGANISM: Aleurain
US-09-120-365-75

Query Match 42.1%; Score 40; DB 3; Length 361;
Best Local Similarity 54.5%; Pred. No. 78;
Matches 12; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

OY 2 RVLVLAALAV---ALAVGDQSN 19
|||:||||| 1:11 1:
DB 5 RVLVLAALAVLAATAVAVAASSSS 26

RESULT 8
US-09-515-039-75
Sequence 75, Application US/09515039
Patent No. 6214599
GENERAL INFORMATION:
APPLICANT: Natori, Shunji
TITLE OF INVENTION: NEW PROTEASE
FILE REFERENCE: 32290-144749
CURRENT APPLICATION NUMBER: US/09/515,039
CURRENT FILING DATE: 2000-03-06
EARLIER APPLICATION NUMBER: JP 9-333 474
EARLIER FILING DATE: 1997-11-18
NUMBER OF SEQ ID NOS: 101
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 75
LENGTH: 361
TYPE: PRT
ORGANISM: Aleurain
US-09-515-039-75

Query Match 42.1%; Score 40; DB 4; Length 361;
Best Local Similarity 54.5%; Pred. No. 78;
Matches 12; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

OY 2 RVLVLAALAV---ALAVGDQSN 19
|||:||||| 1:11 1:
DB 5 RVLVLAALAVLAATAVAVAASSSS 26

RESULT 9
US-08-484-493-12
Sequence 12, Application US/08484493
Patent No. 5728381
GENERAL INFORMATION:
APPLICANT: Wilson, Peter J
APPLICANT: Morris, Charles P
APPLICANT: Anson, Donald S
APPLICANT: Occhiodoro, Teresa
APPLICANT: Bielicki, Julie
APPLICANT: Clements, Peter R
APPLICANT: Hopwood, John J
TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
NUMBER OF SEQUENCES: 15
IDURONATE 2-SULFATASE
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza

CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,493
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 991,973
FILING DATE: 17-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 84162
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 507 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-493-12

Query Match 42.1%; Score 40; DB 1; Length 507;
Best Local Similarity 52.6%; Pred. No. 1.1e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

OY 2 RVLVLAALAVLAAGVQSNL 20
|||:||||| 1:11 1:
DB 5 RSLVLAALAGLAVARPPNI 23

RESULT 10
US-08-484-494-12
Sequence 12, Application US/08484494
Patent No. 5798239
GENERAL INFORMATION:
APPLICANT: Wilson, Peter J
APPLICANT: Morris, Charles P
APPLICANT: Anson, Donald S
APPLICANT: Occhiodoro, Teresa
APPLICANT: Bielicki, Julie
APPLICANT: Clements, Peter R
APPLICANT: Hopwood, John J
TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
NUMBER OF SEQUENCES: 15
IDURONATE 2-SULFATASE
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,494
FILING DATE: 07-JUN-1995

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 991,973
FILING DATE: 17-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8416Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4343
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 507 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-484-494-12

Query Match 42.1%; Score 40; DB 1; Length 507;
Best Local Similarity 52.6%; Pred. No. 1.1e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 RVLVLAVALAVGDSNL 20
| : |||| | : |
Db 5 RSLTALAAGLAVARPPNI 23

RESULT 11
US-08-345-212-12
; Sequence 12, Application US/08345212
; Patent No. 5932211
; GENERAL INFORMATION:
; APPLICANT: Wilson, Peter J
; APPLICANT: Morris, Charles P
; APPLICANT: Anson, Donald S
; APPLICANT: Occhiodoro, Teresa
; APPLICANT: Bielicki, Julie
; APPLICANT: Clements, Peter R
; APPLICANT: Hopwood, John J
; TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
; TITLE OF INVENTION: IDURONATE 2-SULFATASE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/345,212
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 991,973
; FILING DATE: 17-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8416Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR

INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 507 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-345-212-12

Query Match 42.1%; Score 40; DB 2; Length 507;
Best Local Similarity 52.6%; Pred. No. 1.1e+02;
Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;
QY 2 RVLVLAVALAVGDSNL 20
| : |||| | : |
Db 5 RSLTALAAGLAVARPPNI 23

RESULT 12
US-09-249-003-12
; Sequence 12, Application US/09249003
; Patent No. 6153188
; GENERAL INFORMATION:
; APPLICANT: Wilson, Peter J
; APPLICANT: Morris, Charles P
; APPLICANT: Anson, Donald S
; APPLICANT: Occhiodoro, Teresa
; APPLICANT: Bielicki, Julie
; APPLICANT: Clements, Peter R
; APPLICANT: Hopwood, John J
; TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
; TITLE OF INVENTION: IDURONATE 2-SULFATASE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Scully, Scott, Murphy & Presser
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: USA
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/249,003
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/991,973
; FILING DATE: 17-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 8416Z
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 516-742-4343
; TELEFAX: 516-742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 507 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-249-003-12

Query Match 42.1%; Score 40; DB 4; Length 507;
Best Local Similarity 52.6%; Pred. No. 1.1e+02;

Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 RVLTALAVAVAGDQSNL 20
1:|||||11111:
Db 5 RSLTALAVAGLAVARPPNI 23

RESULT 13

US-08-445-586-8

Sequence 8, Application US/08445586

Patent No. 5627050

GENERAL INFORMATION:

APPLICANT: Takeshita, Sunao

APPLICANT: Ito, Toshiaki

APPLICANT: Otawara-Hamamoto, Yoko

APPLICANT: Amann, Egon

TITLE OF INVENTION: Bone-Related Sulfatase-Like Protein and

TITLE OF INVENTION: Process for Its Production

NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

ADDRESSEE: Dunner

STREET: 1300 I Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3314

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/445,586

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/111,887

FILING DATE: 26-AUG-1993

APPLICATION NUMBER: JP 230030/92

FILING DATE: 28-AUG-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 324034/92

FILING DATE: 03-DEC-1992

ATTORNEY/AGENT INFORMATION:

NAME: Forman, David S.

REGISTRATION NUMBER: 33,694

REFERENCE/DOCKET NUMBER: 02481.1322-00000

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-408-4000

TELEFAX: 202-408-4400

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 509 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-445-586-8

Query Match 42.1%; Score 40; DB 1; Length 509;
Best Local Similarity 52.6%; Pred. No. 1.2e+02;

Matches 10; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy 2 RVLTALAVAVAGDQSNL 20
1:|||||11111:
Db 7 RSLTALAVAGLAVARPPNI 25

RESULT 14

US-08-477-396A-4

Sequence 4, Application US/08477396A

Patent No. 5872235

GENERAL INFORMATION:

APPLICANT: Chen, Ian Bo

APPLICANT: Bao, Shideng

APPLICANT: Liu, Yuan

TITLE OF INVENTION: A NOVEL TUMOR MARKER AND NOVEL METHOD OF

TITLE OF INVENTION: ISOLATING SAME

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Weingarten, Schurjin, Gagnebin & Hayes

STREET: Ten Post Office Square

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/477,396A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/146,488

FILING DATE: 29-OCT-1993

APPLICATION NUMBER: US 08/448,388

FILING DATE: 28-MAY-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/12502

FILING DATE: 31-OCT-1994

ATTORNEY/AGENT INFORMATION:

NAME: Heine, Holliday C.

REGISTRATION NUMBER: 34,346

REFERENCE/DOCKET NUMBER: DPCT-333BX

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-2290

TELEFAX: (617) 451-0313

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 777 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-477-396A-4

Query Match 41.1%; Score 39; DB 2; Length 777;
Best Local Similarity 72.7%; Pred. No. 2.7e+02;

Matches 8; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 5 VIALAVAVAG 15
:|||||11111:
Db 7 ILALALALAG 17

RESULT 15

US-08-095-726-14

Sequence 14, Application US/08095726

Patent No. 5530188

GENERAL INFORMATION:

APPLICANT: Auslich, Rodney L

APPLICANT: Brinkhaus, Friedhelm L

APPLICANT: Mukharji, Indrani

APPLICANT: Profitit, John H

APPLICANT: Targier, James G

APPLICANT: Yen, Huel-Che B

TITLE OF INVENTION: Beta-Carotene Biosynthesis in

TITLE OF INVENTION: Genetically Engineered Hosts

NUMBER OF SEQUENCES: 79

CORRESPONDENCE ADDRESS:

ADDRESSEE: Amoco Corp., Patents and Licensing Dept

STREET: 200 E Randolph St

CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60680-0703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,726
FILING DATE: 21-JUL-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/785,566
FILING DATE: 30-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, NO. 553018val B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 3128567180
TELEFAX: 3128564972
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 374 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-095-726-14

Query Match 40.0%; Score 38; DB 1; Length 374;
Best Local Similarity 50.0%; Pred. No. 1.7e+02;
Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
OY 4 LVIALVALAVGDSNLG 21
Db 272 LPLAVLADATADSPRUG 289

Search completed: August 6, 2002, 15:03:05
Job time: 190 sec

FT DISULEFID 124 139 ALTERNATE.
SQ SEQUENCE 180 AA; 19975 MW; C2449BB02A1A80F1 CRC64;

Query Match 46.3%; Score 44; DB 1; Length 180;
Best Local Similarity 55.6%; Pred. No. 17;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 MRVYLALVALVAGDOS 18
1:|||||:|||||1:
DB 1 MKCLLALGLALAGCIGA 18

RESULT 8

LACB_SHEEP STANDARD: PRT: 180 AA.
ID LACB_SHEEP P02757;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Beta-lactoglobulin 1/B, 2/A, and 3/C precursor.
OS Ovis aries (Sheep), and
OS Ovis orientalis musimon (Mouflon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940, 9938;
RN [1]
RP SEQUENCE FROM N.A. (BLG 1 AND 2).
RC SPECIES-SHEEP;
RX MEDLINE=88172489; PubMed=3351935;
RA All S., Clark A.J.;
RT "Characterization of the gene encoding ovine beta-lactoglobulin.
RT Similarity to the genes for retinol binding protein and other
RT secretory proteins.";
RL J. Mol. Biol. 199:415-426(1988).
RN [2]
RP SEQUENCE FROM N.A. (BLG 1).
RC SPECIES-SHEEP;
RX MEDLINE=87049827; PubMed=3096387;
RA Gaye P., Hue-Dehahne D., Mercier J.-C., Soulier S., Vilotte J.-L.,
RA Furet J.-P.;
RT "Ovine beta-lactoglobulin messenger RNA: nucleotide sequence and mRNA
RT levels during functional differentiation of the mammary gland.";
RL Biochimie 68:1097-1107(1986).
RN [3]
RP SEQUENCE FROM N.A. (BLG 1).
RC SPECIES-SHEEP;
RX MEDLINE=89057492; PubMed=3194215;
RA Harris S., All S., Anderson S., Archibald A.L., Clark A.J.;
RT "Complete nucleotide sequence of the genomic ovine beta-lactoglobulin
RT gene.";
RL Nucleic Acids Res. 16:10379-10380(1988).
RN [4]
RP SEQUENCE FROM N.A. (BLG 1 AND 2).
RC SPECIES-SHEEP;
RX MEDLINE=91007276; PubMed=1976573;
RA All S., McClenaghan M., Simons J.P., Clark A.J.;
RT "Characterisation of the alleles encoding ovine beta-lactoglobulins A
RT and B.";
RL Gene 91:201-207(1990).
RN [5]
RP SEQUENCE OF 19-180 (BLG 2).
RC SPECIES-SHEEP;
RX MEDLINE=80219294; PubMed=6155855;
RA Preaux G., Braunltzer G., Kolde H.-J.;
RT "Primary structure of ovine beta-lactoglobulin.";
RL Arch. Int. Physiol. Biochim. 88:B45-B46(1980).
RN [6]
RP SEQUENCE OF 19-180 (BLG 3).
RC SPECIES-SHEEP;
RX MEDLINE=89374823; PubMed=2775495;
RA Erhardt G., Godovac-Zimmermann J., Contl A.;

RT "Isolation and complete primary sequence of a new ovine wild-type
RT beta-lactoglobulin C.";
RL Biol. Chem. Hoppe-Seyler 370:757-762(1989).

RN [7]
RP SEQUENCE OF 19-180 (BLG B).

RC SPECIES-O.O.musimon;
RX MEDLINE=88106996; PubMed=3426802;

RA Godovac-Zimmermann J., Contl A., Napolitano L.;

RT "The complete amino-acid sequence of dimeric beta-lactoglobulin from
RT mouflon (Ovis ammon musimon) milk.";
RL Biol. Chem. Hoppe-Seyler 368:1313-1319(1987).

CC -1- FUNCTION: LACTOGLOBULIN IS THE PRIMARY COMPONENT OF WHEY, IT
CC BINDS RETINOL AND IS PROBABLY INVOLVED IN THE TRANSPORT OF
CC THAT MOLECULE.

CC -1- SUBUNIT: UNDER PHYSIOLOGICAL CONDITIONS BETA-LACTOGLOBULIN EXISTS
CC AS AN EQUILIBRIUM MIXTURE OF MONOMERIC AND DIMERIC FORMS.

CC -1- MISCELLANEOUS: ALTERNATE DISULFIDE BONDS OCCUR IN EQUAL AMOUNTS.

CC -1- SIMILARITY: BELONGS TO THE LIPOCALIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X04520; CAA28204.1; -

DR EMBL: X12817; CAA31305.1; -

DR EMBL: X07004; CAA30059.1; ALT_SEQ.

DR EMBL: X07005; CAA30059.1; JOINED.

DR EMBL: X07006; CAA30059.1; JOINED.

DR EMBL: X07007; CAA30059.1; JOINED.

DR EMBL: X07008; CAA30059.1; JOINED.

DR EMBL: X07009; CAA30059.1; JOINED.

DR EMBL: M32233; AAA31510.1; -

DR EMBL: M32232; AAA31510.1; JOINED.

DR EMBL: M32235; AAA31510.1; JOINED.

DR PIR: A03221; IGSB.

DR PIR: S00132; S00132.

DR PIR: A30011; A30011.

DR PIR: B30011; B30011.

DR PIR: A25136; A25136.

DR PIR: J00748; J00748.

DR PIR: J00749; J00749.

DR PIR: S02136; S02136.

DR PIR: S04955; S04955.

DR HSSP: P02754; 1BSQ.

DR InterPro: IPR002345; Lipocalin.

DR Pfam: PF00061; Lipocalin.1.

DR PRINTS: PR00179; LIPOCALIN.

DR PROSITE: PS00213; LIPOCALIN.1.

KW Milk; Whey; Retinol-binding; Transport; Signal; Lipocalin.

FT SIGNAL 1 18

FT CHAIN 19 180

FT DISULEFID 84 178

FT DISULEFID 124 137

FT DISULEFID 124 139

FT VARIANT 38 38

FT VARIANT 166 166

FT SEQUENCE 180 AA; 19921 MW; BABC8289E575333 CRC64;

Query Match 46.3%; Score 44; DB 1; Length 180;
Best Local Similarity 55.6%; Pred. No. 17;
Matches 10; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 MRVYLALVALVAGDOS 18
1:|||||:|||||1:
DB 1 MKCLLALGLALAGCIGA 18

RESULT 9
 ASN_HUMAN STANDARD: PRT: 629 AA.
 ID ASN_HUMAN STRAND: PRT: 629 AA.
 AC P17405; P17406; Q16857; Q16841; Q13811;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Sphingomyelin phosphodiesterase precursor (EC 3.1.4.12) (Acid
 SMase)
 GN SMPD OR ASM.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 ON NCB1
 NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RX MEDLINE-91217097; PubMed-1840600;
 RA Schuchman E.H., Suchi M., Takahashi T., Sandhoff K., Desnick R.J.;
 RT "Human acid sphingomyelinase. Isolation, nucleotide sequence and
 expression of the full-length and alternatively spliced cDNAs.";
 RL J. Biol. Chem. 266:8531-8539(1991).
 RN [12]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-93183402; PubMed-1292508;
 RA Newrrelia D., Stoffel W.;
 RT "Molecular cloning of the acid sphingomyelinase of the mouse and the
 organization and complete nucleotide sequence of the gene.";
 RL Biol. Chem. Hoppe-Seyler 373:1233-1238(1992).
 RN [13]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92155708; PubMed-1740330;
 RA Schuchman E.H., Levran O., Pereira L.V., Desnick R.J.;
 RT "Structural organization and complete nucleotide sequence of the gene
 encoding human acid sphingomyelinase (SMPD1).";
 RL Genomics 12:197-205(1992).
 RN [14]
 RP SEQUENCE FROM N.A., AND VARIANT ARG-157.
 RX MEDLINE-94012573; PubMed-8407868;
 RA Ida H., Rennett O.M., Eto Y., Chan W.Y.;
 RT "Cloning of a human acid sphingomyelinase cDNA with a new mutation
 that renders the enzyme inactive.";
 RL J. Biochem. 114:15-20(1993).
 RN [15]
 RP SEQUENCE OF 128-629 FROM N.A., PARTIAL SEQUENCE, AND
 ALTERNATIVE SPLICING.
 RC TISSUE-Fibroblast;
 RX MEDLINE-90060003; PubMed-2555181;
 RA Quintern L.E., Schuchman E.H., Levran O., Suchi M., Ferlinz K.,
 Reinke H., Sandhoff K., Desnick R.J.;
 RT "Isolation of cDNA clones encoding human acid sphingomyelinase:
 occurrence of alternatively processed transcripts.";
 RL EMBO J. 8:2469-2473(1989).
 RN [16]
 RP CARBOHYDRATE-BINDING SITES.
 RX MEDLINE-97182640; PubMed-9030779;
 RA Ferlinz K., Hurwitz R., Sandhoff K.;
 RT "Functional characterization of the N-glycosylation sites of human
 acid sphingomyelinase by site-directed mutagenesis.";
 RL Eur. J. Biochem. 243:511-517(1997).
 RN [17]
 RP VARIANT NPD SER-577.
 RX MEDLINE-92028849; PubMed-1718266;
 RA Ferlinz K., Hurwitz R., Sandhoff K.;
 RT "Molecular basis of acid sphingomyelinase deficiency in a patient
 with Niemann-Pick disease type A.";
 RL Biochem. Biophys. Res. Commun. 179:1187-1191(1991).
 RN [18]
 RP VARIANT NPD LEU-496.
 RX MEDLINE-91219449; PubMed-2023926;
 RA Levran O., Desnick R.J., Schuchman E.H.;
 RT "Niemann-Pick disease: a frequent missense mutation in the acid
 sphingomyelinase gene of Ashkenazi Jewish type A and B patients.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3748-3752(1991).
 RN [19]
 RP VARIANT NPD ARG-608 DEL.
 RX MEDLINE-91358737; PubMed-1885770;
 RA Levran O., Desnick R.J., Schuchman E.H.;
 RT "Niemann-Pick type B disease. Identification of a single codon
 deletion in the acid sphingomyelinase gene and genotype/phenotype
 correlations in type A and B patients.";
 RL J. Clin. Invest. 88:806-810(1991).
 RN [10]
 RP VARIANT NPD PRO-302.
 RX MEDLINE-93004773; PubMed-1391960;
 RA Levran O., Desnick R.J., Schuchman E.H.;
 RT "Identification and expression of a common missense mutation (L302P)
 in the acid sphingomyelinase gene of Ashkenazi Jewish type A
 blood 80:2081-2087(1992).
 RN [11]
 RP VARIANT NPD ARG-436.
 RX MEDLINE-93244834; PubMed-1301192;
 RA Takahashi T., Desnick R.J., Takada G., Schuchman E.H.;
 RT "Identification of a missense mutation (S436R) in the acid
 sphingomyelinase gene from a Japanese patient with type B
 Niemann-Pick disease.";
 RL Hum. Mutat. 1:70-71(1992).
 RN [12]
 RP VARIANTS NPD ARG-242; ILE-382 AND SER-383.
 RX MEDLINE-9216934; PubMed-1618760;
 RA Takahashi T., Suchi M., Desnick R.J., Takada G., Schuchman E.H.;
 RT "Identification and expression of five mutations in the human acid
 sphingomyelinase gene causing types A and B Niemann-Pick disease.
 Molecular evidence for genetic heterogeneity in the neuropathic and
 non-neuropathic forms.";
 RL J. Biol. Chem. 267:12552-12558(1992).
 RN [13]
 RP VARIANT NPD GLY-391.
 RX MEDLINE-94328611; PubMed-8051942;
 RA Speil W., Bart G., Vanier M.T., Christomanou H., Baldissera I.,
 Reichensteiner E., Paschke E.;
 RT "A family with visceral course of Niemann-Pick disease, macular halo
 syndrome and low sphingomyelin degradation rate.";
 RL J. Inher. Metab. Dis. 17:93-103(1994).
 RN [14]
 RP VARIANT NPD THR-389.
 RX MEDLINE-96287387; PubMed-8680412;
 RA Schuchman E.H.;
 RT "Two new mutations in the acid sphingomyelinase gene causing type A
 Niemann-Pick disease: N389T and R441X.";
 RL Hum. Mutat. 6:352-354(1995).
 RN [15]
 RP VARIANT NPD CYS-446.
 RX MEDLINE-96274768; PubMed-8693491;
 RA Takahashi T., Suchi M., Sato M., Ten S.B., Sakuragawa N.,
 Desnick R.J., Schuchman E.H., Takada G.;
 RT "Identification and expression of a missense mutation (Y446C) in the
 acid sphingomyelinase gene from a Japanese patient with type A
 Niemann-Pick disease.";
 RL Tohoku J. Exp. Med. 177:117-123(1995).
 RN [16]
 RP VARIANT NPD GLN-246.
 RX MEDLINE-96263741; PubMed-8664904;
 RA Ida H., Rennett O.M., Maekawa K., Eto Y.;
 RT "Identification of three novel mutations in the acid
 sphingomyelinase gene of Japanese patients with Niemann-Pick disease
 type A and B.";
 RL Hum. Mutat. 7:65-67(1996).
 CC -1- FUNCTION: CONVERTS SPHINGOMYELIN TO CERAMIDE. ASM ALSO HAS
 PHOSPHOLIPASE C ACTIVITIES TOWARD 1,2-DIACYLGLYCEROLPHOSPHOCHOLINE
 AND 1,2-DIACYLGLYCEROLPHOSPHOGLYCEROL.
 CC -1- CATALYTIC ACTIVITY: Sphingomyelin + H(2)O - N-acylsphingosine +
 choline phosphate.
 CC -1- SUBUNIT: MONOMER.

CC -1- SUBCELLULAR LOCATION: Lysosomal.
 CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; ASM-1 (SHOWN HERE)
 CC (FREQUENCY 90%), ASM-2 (FREQUENCY 10%) AND ASM-3 (FREQUENCY <1%);
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. ONLY ASM-1 ENCODES A
 CC CATALYTICALLY ACTIVE ENZYME.
 CC -1- DISEASE: DEFECTS IN SMPD1 ARE A CAUSE OF THE NIEMANN-PICK DISEASE
 CC (NPD). SPHINGOMELIN AND OTHER METABOLICALLY RELATED LIPIDS
 CC ACCUMULATE IN THE LYSOSOMES. PATIENTS MAY SHOW XANTHOMAS,
 CC PIGMENTATION, HEPATOSPLENOMEGALY, LYMPHADENOPATHY AND MENTAL
 CC RETARDATION. THERE ARE TWO MAIN TYPES: TYPE A, IS A RAPIDLY
 CC PROGRESSIVE AND FATAL NEURODEGENERATIVE DISEASE OF INFANCY. TYPE
 CC B, HAS LITTLE IF ANY NEUROLOGIC INVOLVEMENT AND MAY SURVIVE INTO
 CC ADULTHOOD. NPD OCCURS MORE FREQUENTLY AMONG INDIVIDUALS OF
 CC ASKENAZI JEWISH ANCESTRY THAN IN THE GENERAL POPULATION.
 CC -1- MISCELLANEOUS: THERE ARE TWO TYPES OF SPHINGOMELINASES: ASM
 CC (ACID), AND NSM (NEUTRAL).
 CC -1- SIMILARITY: BELONGS TO THE ACID SPHINGOMELINASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 SAPOSIN B-TYPE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; M59916; AAA58377.1; -
 CC DR EMBL; M59917; AAA58378.1; -
 CC DR EMBL; X63600; CAA45145.1; -
 CC DR EMBL; X52678; CAA36901.1; -
 CC DR EMBL; X52679; CAA36902.1; -
 CC DR EMBL; M81780; AAA75008.1; -
 CC DR EMBL; M81780; AAA75009.1; -
 CC DR EMBL; X59960; CAA42584.1; -
 CC DR PIR; A39825; A39825.
 CC DR PIR; S06957; S06957.
 CC DR PIR; S06958; S06958.
 CC DR PIR; S27009; S27009.
 CC DR MIM; 257200; -
 CC DR InterPro: IPR000004; Sarp.
 CC DR InterPro: IPR000934; Ser_thr_phosphatase.
 CC DR SMART; SM00118; SAPB; 1.
 CC KM Hydrolase; Glycosidase; Lysosome; Glycoprotein; Alternative splicing;
 CC Signal; Disease mutation.
 CC FT SIGNAL 1 46
 CC FT CHAIN 47 629 SPHINGOMELIN PHOSPHODIESTERASE.
 CC FT DOMAIN 85 169 SAPOSIN-LIKE TYPE B.
 CC FT CARBOHD 86 86 N-LINKED (GLCNAC. . .).
 CC FT CARBOHD 175 175 N-LINKED (GLCNAC. . .).
 CC FT CARBOHD 335 335 N-LINKED (GLCNAC. . .).
 CC FT CARBOHD 395 395 N-LINKED (GLCNAC. . .).
 CC FT CARBOHD 520 520 N-LINKED (GLCNAC. . .).
 CC FT VARSPLIC 363 374 IGFYALSPYPG -> YLSSVTEQCKR (IN ISOFORM
 CC FT VARSPLIC 375 418 ASM-2).
 CC FT VARSPLIC 363 418 MISSING (IN ISOFORM ASM-2).
 CC FT VARSPLIC 363 418 MISSING (IN ISOFORM ASM-3).
 CC
 CC Query Match 46.3%; Score 44; DB 1; Length 629;
 CC Best Local Similarity 71.4%; Pred. No. 50;
 CC Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 CC
 CC Db 36 VIALALVALAVGD 16
 CC 36 VIALALVALALALSD 49
 CC
 CC RESULT 10
 CC CASB_CAPHI STANDARD; PRT; 222 AA.
 CC AC P33048;
 CC DT 01-OCT-1993 (Rel. 27, Created)
 CC DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Beta casein precursor.
 GN CSN2.
 OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OC NCBI_Taxid=9925;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SANEN; TISSUE=Blood;
 RX MEDLINE=93077039; PubMed=1446822;
 RA Roberts B., Dittullo P., Vitale J., Behir K., Gordon K.;
 FT Cloning of the goat beta casein-encoding gene and expression in
 FT transgenic mice";
 RT Gene 121:255-262(1992).
 CC -1- FUNCTION: IMPORTANT ROLE IN DETERMINATION OF THE SURFACE
 CC PROPERTIES OF THE CASEIN MICELLES.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: MAMMARY GLAND; MILK.
 CC -1- SIMILARITY: BELONGS TO THE BETA-CASEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; M90561; AAA30906.1; -
 CC DR EMBL; M90556; AAA30906.1; JOINED.
 CC DR EMBL; M90557; AAA30906.1; JOINED.
 CC DR EMBL; M90558; AAA30906.1; JOINED.
 CC DR EMBL; M90560; AAA30906.1; JOINED.
 CC DR PIR; JCI384; JCI384.
 CC DR InterPro: IPR001588; Casein.
 CC DR Pfam; PF00363; caseins; 1.
 CC DR PROSITE; PS00306; CASEIN.ALPHA.BETA.FALSE.NEG.
 CC KM Milk; Phosphorylation; Glycoprotein; Signal.
 CC FT SIGNAL 1 15
 CC FT CHAIN 16 222 BETA CASEIN.
 CC FT MOD_RES 30 30 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 32 32 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 33 33 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 34 34 PHOSPHORYLATION (BY SIMILARITY).
 CC FT MOD_RES 50 50 PHOSPHORYLATION (BY SIMILARITY).
 CC SQ SEQUENCE 222 AA; 24865 MW; 96AE17746A01CD05 CRC64;
 CC
 CC Query Match 45.8%; Score 43.5; DB 1; Length 222;
 CC Best Local Similarity 52.4%; Pred. No. 25;
 CC Matches 11; Conservative 4; Mismatches 5; Indels 1; Gaps 1;
 CC
 CC Db 1 MRYVIALAVAVG-DQSNL 20
 CC 1 MKVLIILACVIALAVARQEEEL 21
 CC
 CC RESULT 11
 CC AP11_ONCMY STANDARD; PRT; 262 AA.
 CC ID AP11_ONCMY
 CC AC 057523;
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
 CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
 CC DE Apolipoprotein A-I-1 precursor (APOA-I-1).
 CC OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
 CC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
 CC NCBI_Taxid=8022;
 CC RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-SHASTA; TISSUE-LIVER;
 RA MEDLINE-92235558; PubMed-1569377;
 RA Delucue G.P., Sun J.M., Davie J.R.;
 RT "Expression of rainbow trout apolipoprotein A-I genes in liver and
 RT hepatocellular carcinoma.";
 RL J. Lipid Res. 33:251-262(1992).
 CC -1- FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF
 CC CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING
 CC THE LECITHIN CHOLESTEROL ACYLTRANSFERASE (LCAT) (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Extracellular (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF04218; AAB96972.1; -
 DR InterPro: IPR000074; Apolipoprotein.
 DR Pfam: PF01442; Apolipoprotein; 1.
 KW Plasma; Lipid transport; HDL; Cholesterol metabolism; Repeat; Signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT PROPEP 19 23
 FT CHAIN 24 262 APOLOPROTEIN A-I-1.
 FT DOMAIN 32 63 3 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 64 262 10 X APPROXIMATE TANDEM REPEATS.
 FT REPEAT 64 85 1.
 FT REPEAT 87 107 2.
 FT REPEAT 108 118 3 (HALF-LENGTH).
 FT REPEAT 119 140 4.
 FT REPEAT 141 162 5.
 FT REPEAT 163 184 6.
 FT REPEAT 185 206 7.
 FT REPEAT 207 228 8.
 FT REPEAT 229 239 9 (HALF-LENGTH).
 FT REPEAT 240 262 10.
 SO SEQUENCE 262 AA; 29704 MW; B4C311426BA3DB1A CRC64;

Query Match 45.3%; Score 43; DB 1; Length 262;
 Best Local Similarity 50.0%; Pred. No. 33;
 Matches 9; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 MRVVLALAVAVAGDS 18
 I: | | | | | | | | | |
 DB 1 MKFLALATITLAAAGTGA 18

RESULT 12
 ID GBLI_CHITH STANDARD; PRT; 158 AA.
 AC P02221;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Globin CTT-I/CTT-IA precursor (Erythrocytorin).
 GN CTT-1.
 OS Chironomus thummi thummi (Midge).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
 OC Chironomidae; Chironomidae; Chironominae; Chironomus.
 NCBI_TaxID=7155;
 RN [1]
 RP SEQUENCE FROM N.A. (CTT-I).
 RX MEDLINE-95287399; PubMed-7769612;
 RA Kao W.Y., Hankeln T., Schmidt E.R., Bergtrom G.;
 RT "Sequence and evolution of the gene for the monomeric globin I and
 RT its linkage to genes coding for dimeric globins in the insect

RT Chironomus thummi.";
 RL J. Mol. Evol. 40:354-361(1995).
 RN [2]
 RP SEQUENCE OF 16-158 (CTT-1).
 RX MEDLINE-80202931; PubMed-7380386;
 RA Kleinschmidt T., von der Mark-Neuwirth H., Braunitzer G.;
 RT "Hemoglobin, xxxi. Analysis of the primary structure of the monomeric
 RT hemoglobin CTT I (erythrocytorin) of Chironomus thummi thummi,
 RT Diptera.";
 RL Hoppe-Seyler's Z. Physiol. Chem. 361:401-411(1980).
 RN [3]
 RP SEQUENCE OF 16-158 (CTT-1A).
 RX MEDLINE-83236193; PubMed-6862374;
 RA Goodman M., Braunitzer G., Kleinschmidt T., Aschauer H.;
 RT "The analysis of a protein-polymorphism. Evolution of monomeric and
 RT homodimeric haemoglobins (erythrocytorins) of Chironomus thummi
 RT thummi (Insecta, Diptera).";
 RL Hoppe-Seyler's Z. Physiol. Chem. 364:205-217(1983).
 RN [4]
 RP SEQUENCE OF 83-158 FROM N.A. (CTT-1A).
 RX MEDLINE-91276274; PubMed-2055487;
 RA Saffarini D.A., Trewitt P.M., Luhm R.A., Bergtrom G.;
 RT "Differential regulation of insect globin and actin mRNAs during
 RT larval development in Chironomus thummi.";
 RL Gene 101:215-222(1991).
 CC -1- SUBUNIT: MONOMER.
 CC -1- MISCELLANEOUS: THERE ARE AT LEAST 12 DIFFERENT COMPONENTS IN MIDGE
 CC GLOBIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U14627; AAA80189.1; -
 DR EMBL: U14628; AAA80190.1; -
 DR EMBL: U14629; AAA80191.1; -
 DR EMBL: M57410; AAA62727.1; -
 DR PIR: A02543; GCICEL.
 DR PIR: A91719; GGCICL.
 DR HSSP: P02229; IECA.
 DR InterPro: IPR000971; Globin.
 DR Pfam: PF00042; globin; 1.
 DR PROSITE: PS01033; GLOBIN; 1.
 KW Heme; Oxygen transport; Transport; Signal.
 FT SIGNAL 1 15
 FT CHAIN 16 158 GLOBIN CTT-I.
 FT METAL 74 74 IRON (HEME DISTAL LIGAND)
 FT METAL 109 109 IRON (HEME PROXIMAL LIGAND)
 FT METAL 109 109 (BY SIMILARITY).
 FT VARIANT 113 113 A -> T (IN CTT-1A).
 SO SEQUENCE 158 AA; 17011 MW; 9639E8A38908B8AB CRC64;

Query Match 44.2%; Score 42; DB 1; Length 158;
 Best Local Similarity 52.4%; Pred. No. 30;
 Matches 11; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

OY 1 MRVVLALAVAVAVAGDQ 17
 I: | | | | | | | | | |
 DB 1 MKFLIILALCVAAAMAGPSGQ 21

RESULT 13
 ID PBP_MANSE STANDARD; PRT; 168 AA.
 AC P18959;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)

DE 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Pheromone-binding protein precursor (PBP).
 OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
 OS Eukaryota; Metazoa; Arthropoda; Insecta; Hexapoda; Insecta;
 CC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 CC Sphingodea; Sphingidae; Sphinginae; Manduca.
 NCBI_TaxID=7130;
 RX [1]
 RX MEDLINE=9071794; PubMed=3200861.
 RA Györeyi T.K., Rody-Siemkovits A.J., Lerner M.R.;
 RT "Characterization and cDNA cloning of the pheromone-binding protein
 from the tobacco hornworm, Manduca sexta: a tissue-specific
 developmentally regulated protein.";
 RT Proc. Natl. Acad. Sci. U.S.A. 85:9851-9855(1988).
 RN [2]
 RN MEDLINE=91186129; PubMed=2010751;
 RX Vogt R.G., Prestwich G.D., Lerner M.R.;
 RT "Odorant-binding protein subfamilies associate with distinct classes
 of olfactory receptor neurons in insects.";
 RT J. Neurobiol. 22:74-84(1991).
 RL J. Neurobiol. 22:74-84(1991).
 CC FUNCTION: THIS MAJOR SOLUBLE PROTEIN IN OLFACTORY SENSILLA OF MALE
 MOTHS MIGHT SERVE TO SOLUBILIZE THE EXTREMELY HYDROPHOBIC
 PHEROMONE MOLECULES AND TO TRANSPORT PHEROMONE THROUGH THE AQUEOUS
 LYMPH TO RECEPTORS LOCATED ON OLFACTORY CILIA. PBP IS ALSO FOUND
 IN SENSILLA FROM FEMALE M. SEXTA ANTENNAE.
 CC SUBUNIT: HOMODIMER (PROBABLE).
 CC TISSUE SPECIFICITY: ANTENNA.
 CC DEVELOPMENTAL STAGE: ITS SYNTHESIS OCCURS AROUND THE TIME OF
 ECLOSION.
 CC SIMILARITY: BELONGS TO THE PBP/GOBP FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: M21797; AAA29325.1;
 DR EMBL: M21798; AAA29326.1;
 DR PIR: A31770; A31770.
 DR InterPro: IPR000746; PBP_GOBP.
 DR Pfam: PF01395; PBP_GOBP.1.
 DR PRINTS: PR00484; PBP_GOBP.
 KW Pheromone response; Transport; Signal.
 FT STGM 1 26
 FT CHAIN 27 168 PHEROMONE-BINDING PROTEIN.
 FT VARIANT 29 29 D -> E.
 FT VARIANT 120 120 I -> V.
 FT SEQUENCE 168 AA; 18516 MW; 0FC1E18D1908ADF1 CRC64;
 SQ
 Query Match 44.2%; Score 42; DB 1; Length 168;
 Best Local Similarity 56.2%; Pred. No. 31;
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RX [1]
 RX MEDLINE=9071794; PubMed=3200861.
 RA Györeyi T.K., Rody-Siemkovits A.J., Lerner M.R.;
 RT "Characterization and cDNA cloning of the pheromone-binding protein
 from the tobacco hornworm, Manduca sexta: a tissue-specific
 developmentally regulated protein.";
 RT Proc. Natl. Acad. Sci. U.S.A. 85:9851-9855(1988).
 RN [2]
 RN MEDLINE=91186129; PubMed=2010751;
 RX Vogt R.G., Prestwich G.D., Lerner M.R.;
 RT "Odorant-binding protein subfamilies associate with distinct classes
 of olfactory receptor neurons in insects.";
 RT J. Neurobiol. 22:74-84(1991).
 RL J. Neurobiol. 22:74-84(1991).
 CC FUNCTION: THIS MAJOR SOLUBLE PROTEIN IN OLFACTORY SENSILLA OF MALE
 MOTHS MIGHT SERVE TO SOLUBILIZE THE EXTREMELY HYDROPHOBIC
 PHEROMONE MOLECULES AND TO TRANSPORT PHEROMONE THROUGH THE AQUEOUS
 LYMPH TO RECEPTORS LOCATED ON OLFACTORY CILIA. PBP IS ALSO FOUND
 IN SENSILLA FROM FEMALE M. SEXTA ANTENNAE.
 CC SUBUNIT: HOMODIMER (PROBABLE).
 CC TISSUE SPECIFICITY: ANTENNA.
 CC DEVELOPMENTAL STAGE: ITS SYNTHESIS OCCURS AROUND THE TIME OF
 ECLOSION.
 CC SIMILARITY: BELONGS TO THE PBP/GOBP FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL: M21797; AAA29325.1;
 DR EMBL: M21798; AAA29326.1;
 DR PIR: A31770; A31770.
 DR InterPro: IPR000746; PBP_GOBP.
 DR Pfam: PF01395; PBP_GOBP.1.
 DR PRINTS: PR00484; PBP_GOBP.
 KW Pheromone response; Transport; Signal.
 FT STGM 1 26
 FT CHAIN 27 168 PHEROMONE-BINDING PROTEIN.
 FT VARIANT 29 29 D -> E.
 FT VARIANT 120 120 I -> V.
 FT SEQUENCE 168 AA; 18516 MW; 0FC1E18D1908ADF1 CRC64;
 SQ
 Query Match 44.2%; Score 42; DB 1; Length 168;
 Best Local Similarity 56.2%; Pred. No. 31;
 Matches 9; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

FT	CARBOHD	941	941	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	945	945	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	954	954	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	1004	1004	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	1019	1019	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	1083	1083	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	1142	1142	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	1179	1179	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	1257	1257	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	1292	1292	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	1342	1342	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	1361	1361	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	1366	1366	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	1390	1390	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	1577	1577	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
FT	CARBOHD	1655	1655	N-LINKED	(GLC.NAC. . .)	(POTENTIAL)
50	SEQUENCE	1687	18605	MM: 4965BB9D5FB928F	CRC64:	

```

Query March 13, 1998      61.1%; Score 58; DB 1; Length 1687;
Best Local Similarity 76.5%; Pred. No. 1,2;
Matches 13; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Cy      1 MRV1VALVALVALVAGDQ 17
        ||||| 111 1:1
Db      1 MRV1VALVALVALVAGNQ 17

RESULT 2
APAL_BRARE
ID      APAL_BRARE      STANDARD:      PRT:      262 AA.
CY      042363;
DR      15-DEC-1998 (Rel. 37, Created)
DR      15-DEC-1998 (Rel. 37, Last sequence update)
DR      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Apolipoprotein A-I precursor (Apo-A1).
GN      APOA.
OS      Brachydanio rerio (Zebrafish) (Zebra danio).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC      Cypriniformes; Cyprinidae; Danio.
OX      NCBI_TaxID=7955;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Embryo;
RX      MEDLINE=97385153; PubMed=9238027;
RA      Babin P.J., Thisse C., Durliat M., Andre M., Akimenko M.-A.,
RT      Thisse B.;
RT      "Both apolipoprotein E and A-I genes are present in a nonmammalian
RT      vertebrate and are highly expressed during embryonic development.",
RL      Proc. Natl. Acad. Sci. U.S.A. 94:8622-8627(1997).
CC      -1- FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF
CC      CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING
CC      CHOLESTEROL EFFLUX FROM TISSUES AND BY ACTING AS A COFACTOR FOR
CC      THE LECTININ CHOLESTEROL ACETYLTRANSFERASE (LCAT) (BY SIMILARITY).
CC      -1- SUBCELLULAR LOCATION: Extracellular.
CC      -1- DEVELOPMENTAL STAGE: HIGHLY EXPRESSED IN THE YOLK SYNCYTIAL LAYER
CC      DURING EMBRYONIC (STARTING AT THE GASTRULA STAGE) AND EARLY LARVAL
CC      DEVELOPMENT". AN EXTRAMEMBRYONIC STRUCTURE IMPLICATED IN EMBRYONIC
CC      AND LARVAL NUTRITION.
CC      -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: Y13653; CAAT74004.1; -.
CC      ZFIN: ZDB-GENE-990415-14; apoA.
CC      InterPro: IPR000074; Apolipoprotein.
DR

```

[illegible]

```

Query Match          52.6%:  Score 50;  DB 1;  Length 262;
Best Local Similarity 50.0%:  Pred. NO. 3.5;
Matches 10;  Conservative 5;  Mismatches 5;  Indels 0;  Gaps 0;

QY      1  MKVVLALVALVAVGDOSNL 20
      1:  ||| : ||:|:|:|
DB      1  MKFVALTLTLALGSOANL 20

RESULT 3
APRL_SALSA
ID  APAL_SALSA      STANDARD:      PRT;      258 AA.
AC  P27007:
DR  01-AUG-1992 (Rel. 23, Created)
DT  01-AUG-1992 (Rel. 23, Last sequence update)
DT  30-MAY-2000 (Rel. 39, Last annotation update)
DE  Apolipoprotein A-I precursor (Apo-AI).
OS  Salmo salar (Atlantic salmon).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC  Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX  NCBI_TaxId=8030;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Liver;
RA  MEDLINE=92009208; PubMed=1916286;
RA  Powell R., Higgins D.G., Wolff J., Byrnes L., Stack M., Sharp P.M.,
RA  Cannon F.;
RT  "The salmon gene encoding apolipoprotein A-I: cDNA sequence, tissue
RT  expression and evolution."
RL  Gene 104:155-161(1991).

CC  -1- FUNCTION: APOA-1 PARTICIPATES IN THE REVERSE TRANSPORT OF
CC  CHOLESTEROL FROM TISSUES TO THE LIVER FOR EXCRETION BY PROMOTING
CC  CHOLESTEROL EFFLUX FROM TISSUES AND BY ACTING AS A COFACTOR FOR
CC  THE LECITHIN CHOLESTEROL ACYLTRANSFERASE (LCAT).
CC  -1- SUBCELLULAR LOCATION: Extracellular.
CC  -1- TISSUE SPECIFICITY: MAJOR PROTEIN OF PLASMA HDL, ALSO FOUND IN
CC  CHLOMICRONS. EXPRESSED IN LIVER, INTESTINE, AND MUSCLE.
CC  -1- SIMILARITY: BELONGS TO THE APOA1 / APOA4 / APOE FAMILY.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  EMBL: X52237; CA36482.1; -
DR  PIR: JH0472; JH0472.
DR  PIR: S26810; S26810.
DR  InterPro: IPR000074; Apolipoprotein.
DR  Pfam: PF01442; Apolipoprotein; 1.

```

[illegible]

```

CC modified and this statement is not removed. Usage by and for commercial
CC entitled requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----CC
DR EMBL; L33796; AAA58785.1; -
DR EMBL; AE004338; -; NOT_ANNOTATED_CDS.
DR TIGR; WC2733; -
DR InterPro; IPR001775; Bac_GSPD.
DR InterPro; IPR000016; Bac_GSPprotein.
DR Pfam; PF00263; GSP11_11; 1.
DR PRINTS; PRO0811; BACTERIALGSPD.
DR PROSITE; PS00875; T2SP_D; 1.
RW Transport; Outer membrane; Signal; Complete proteome.
FT SIGNAL 1 24 POTENTIAL
FT CHAIN 25 674 GENERAL SECRETION PATHWAY PROTEIN D.
FT FT CONFLICT 89 89 V -> A (IN REF. 1).
FT FT CONFLICT 144 144 R -> P (IN REF. 1).
SQ SEQUENCE 674 AA; 73469 MW; 3077B891A59E6223 CRC64;

Query Match 49.5%; Score 47; DB 1; Length 674;
Best Local Similarity 55.0%; Pred. No. 20;
Matches 11; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Oy 2 RVLVLAALVALAGDOSNLG 21
      :||: || | :| || |||
Db 343 QVLEALIVEMAEQDGNIHG 362

RESULT 5
GSPD_AERHY ID GSPD_AERHY STANDARD: PRT; 678 AA.
AC P31780;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE General secretion pathway protein D precursor.
GN Exed.
OS Aeromonas hydrophila.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
OX NCBI_TaxID=644;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-AH65;
RC Howard S.P.;
RL Submitted (SEP-1994) to the EMBL/Genbank/DBJ databases.
RM [2]
RP SEQUENCE OF 408-678 FROM N.A.
RC STRAIN-AH65;
RX MEDLINE-92349963; PubMed-1640836;
RA Jiang B., Howard S.P.;
RT "The Aeromonas hydrophila exeE gene, required both for protein
RT secretion and normal outer membrane biogenesis, is a member of a
RT general secretion pathway";
MOL. Microbiol. 6:1351-1361(1992)..
CC -I- FUNCTION: INVOLVED IN A GENERAL SECRETION PATHWAY (GSP) FOR THE
CC EXPORT OF PROTEINS.
CC -I- SIMILARITY: BELONGS TO THE EXED/CUTD/PUD/XPSD FAMILY.
-----CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----CC
DR EMBL; X66504; CAA47124.1; -
DR PIR; S22668; S22668.
DR InterPro; IPR001775; Bac_GSPD.
DR InterPro; IPR000016; Bac_GSPprotein.
```


XX Isolated nucleic acid for assaying for heterologous gene expression,
PT detecting presence of compound that binds to estrogen receptor or
PT producing desired protein from host cell comprises nucleotide sequence
PT encoding secretory signal sequence
XX
PS Claim 1: Page 36; 73pp; English.
XX
CC This sequence is that of a piscine, Oreochromis aureus, vitellogenin
CC secretory sequence (Vtgs). This and variants that comprise conservative
CC replacements that retain the biological activities of directing secretion
CC of a fusion protein from a cell and cleavage of the secretory signal
CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be
CC fused to either a reporter protein or a lipopolysaccharide-binding
CC protein coding sequence. The isolated nucleic acid is useful for assaying
CC for heterologous gene expression, detecting the presence of a compound
CC that binds to an estrogen receptor in a sample or producing a desired
CC protein from a host cell. It can also be used in a method for obtaining
CC systemic circulation of a desired protein in a transgenic or chimeric
CC host organism.
CC
XX Sequence 21 AA;
SQ

Query Match 100.0%; Score 95; DB 21; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MRVLALAVAVAVGDSNG 21
| | | | | | | | | | | | | | | | | | | | | |
Db 1 mrvlvalavavavdgdsnlg 21

RESULT 2
AA92780
ID AA92780 standard; Protein; 34 AA.
XX
AC AA92780;
XX
DT 29-AUG-2000 (first entry)
XX
DE Vtgs-CAT fusion protein (partial).
XX
DE Vtgs: vitellogenin; secretory signal sequence; gene expression;
KM oestrogen receptor binding protein; systemic circulation; CAT.
XX
OS Chimeric - Oreochromis aureus.
OS Chimeric - Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= "secretory_signal_sequence"
FT /note= "O. aureus"
FT Cleavage-site 15..16
FT Protein 22..34
FT /note= "CAT N-terminal"
XX
PN WO200026366-A1.
XX
PD 11-MAY-2000.
XX
PF 29-OCT-1999; 99WO-SG00108.
XX
PR 30-OCT-1998; 98US-0106426.
PR 26-OCT-1999; 99US-0426776.
XX
PA (UYSI-) UNIV SINGAPORE NAT.
PA (LAMT/) LAM T J.
XX
PI Ding JL, Tan NS, Ho B;
XX
DR MPI; 2000-365615/31.
DR N-PDB; AAA28499.
DR

XX Isolated nucleic acid for assaying for heterologous gene expression,
PT detecting presence of compound that binds to estrogen receptor or
PT producing desired protein from host cell comprises nucleotide sequence
PT encoding secretory signal sequence
XX
PS Example 2: Fig 5B; 73pp; English.
XX
CC A reporter CAT system that uses the piscine, Oreochromis aureus,
CC vitellogenin secretory sequence (Vtgs), designated psp-VtgsCAT was
CC constructed. AAA28499 and AAA28722 comprise the 5'- and 3'-ends of the
CC construct insert. Vtgs and variants that comprise conservative
CC replacements that retain the biological activities of directing
CC secretion of a fusion protein from a cell and cleavage of the secretory
CC signal sequence from the fusion protein, are new. DNA encoding the Vtgs
CC can be fused to either a reporter protein or a
CC lipopolysaccharide-binding protein coding sequence. The isolated nucleic
CC acid is useful for assaying for heterologous gene expression, detecting
CC the presence of a compound that binds to an estrogen receptor in a
CC sample or producing a desired protein from a host cell. It can also be
CC used in a method for obtaining systemic circulation of a desired protein
CC in a transgenic or chimeric host organism.
CC
XX Sequence 34 AA;
SQ

Query Match 100.0%; Score 95; DB 21; Length 34;
Best Local Similarity 100.0%; Pred. No. 2.5e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MRVLALAVAVAVGDSNG 21
| | | | | | | | | | | | | | | | | | | | | |
Db 1 mrvlvalavavavdgdsnlg 21

RESULT 3
AA92782
ID AA92782 standard; Protein; 38 AA.
XX
AC AA92782;
XX
DT 29-AUG-2000 (first entry)
XX
DE Vtgs-EGFP fusion protein (partial).
XX
DE Vtgs: vitellogenin; secretory signal sequence; gene expression;
KM oestrogen receptor binding protein; systemic circulation; EGFP.
XX
OS Chimeric - Oreochromis aureus.
OS Chimeric - Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= Vtgs
FT Cleavage-site 15..16
FT Protein 28..38
FT /label= EGFP
XX
PN WO200026366-A1.
XX
PD 11-MAY-2000.
XX
PF 29-OCT-1999; 99WO-SG00108.
XX
PR 30-OCT-1998; 98US-0106426.
PR 26-OCT-1999; 99US-0426776.
XX
PA (UYSI-) UNIV SINGAPORE NAT.
PA (LAMT/) LAM T J.
XX
PI Ding JL, Tan NS, Ho B;
XX
DR MPI; 2000-365615/31.
DR

DR N-PSDB: AAA28502.

XX Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence

XX Example 3; Fig 8B; 73pp; English.

XX A reporter GFP system that uses the piscine, Oreochromis aureus,
 CC vitellogenin secretory sequence (Vtgs), designated pVtgsGFP was
 CC constructed. Vtgs and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an oestrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.

XX Sequence 38 AA:

Query Match 100.0%; Score 95; DB 21; Length 38;

Best Local Similarity 100.0%; Pred. No. 2.8e-07;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MRVLVLAVALAVGDSNMG 21
 |||||
 Db 1 mrvlvlaalavavgdsnlg 21

RESULT 4

AA92779 AAY92779 standard; Protein; 51 AA.

AC AAY92779;

DT 29-AUG-2000 (first entry)

DE Vtgs-CrFCS fusion protein.

XX Vtgs: vitellogenin; secretory signal sequence; gene expression;
 KW oestrogen receptor binding protein; systemic circulation; factor C.

XX Chimeric - Oreochromis aureus.

OS Chimeric - Carcinocorpius rotundicauda.

XX Key Location/Qualifiers

FT Peptide 1..21 /label= secretory_signal_sequence

FT Cleavage-site 15..16 /note= "from O. aureus"

FT Domain 22..51 /label= lps-binding_domain

FT /note= "from C. rotundicauda Factor C"

PN WO200026366-A1.

PD 11-MAY-2000.

PF 29-OCT-1999; 99WO-SG00108.

PR 30-OCT-1998; 98US-0106426.

PR 26-OCT-1999; 99US-0426776.

XX (UYSI-) UNIV SINGAPORE NAT.
 PA (LAMT/) LAM T J.
 XX Ding JL, Tan NS, Ho B;
 PI

XX WPI: 2000-365615/31.
 DR N-PSDB: AAA28496.

XX Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence

XX Example 1; Fig 2A; 73pp; English.

XX The piscine, Oreochromis aureus, vitellogenin secretory sequence (Vtgs)
 CC was fused upstream to the EcoRI-SalI cDNA fragment encoding the
 CC lipopolysaccharide-binding domain of Carcinocorpius rotundicauda Factor
 CC CrFCS for expression and secretion of recombinant ES protein from
 CC Drosophila cells. Vtgs and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Vtgs can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an oestrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.

XX Sequence 51 AA:

Query Match 100.0%; Score 95; DB 21; Length 51;

Best Local Similarity 100.0%; Pred. No. 3.8e-07;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MRVLVLAVALAVGDSNMG 21
 |||||
 Db 1 mrvlvlaalavavgdsnlg 21

RESULT 5

AA92783 AAY92783 standard; Protein; 51 AA.

AC AAY92783;

DT 29-AUG-2000 (first entry)

DE Vtgs-beta-lactamase fusion protein (partial).

XX Vtgs: vitellogenin; secretory signal sequence; gene expression;
 KW oestrogen receptor binding protein; systemic circulation;

XX beta-lactamase.

OS Chimeric - Oreochromis aureus.

OS Chimeric - Synthetic.

XX Key Location/Qualifiers

FT Peptide 1..21 /label= Vtgs

FT Cleavage-site 15..16 /label= Protein

FT /note= "beta-lactamase mature protein"

PN WO200026366-A1.

PD 11-MAY-2000.

PF 29-OCT-1999; 99WO-SG00108.

PR 30-OCT-1998; 98US-0106426.

PR 26-OCT-1999; 99US-0426776.

XX (UYSI-) UNIV SINGAPORE NAT.
 PA

PA (LMT/) LAM T J.
 XX
 PI Ding JL, Tan NS, Ho B;
 XX
 DR WPI; 2000-365615/31.
 DR N-PSDB; AAA28507.
 XX
 PT Isolated nucleic acid for assaying for heterologous gene expression,
 PT detecting presence of compound that binds to estrogen receptor or
 PT producing desired protein from host cell comprises nucleotide sequence
 PT encoding secretory signal sequence
 XX
 PS Example 6; Fig 14A; 73pp; English.
 XX
 CC A reporter beta-lactamase system that uses the *Oreochromis aureus*,
 CC vitellogenin secretory sequence (Ytgss), designated pBADYgbiactana was
 CC constructed. Ytgss and variants that comprise conservative
 CC replacements that retain the biological activities of directing secretion
 CC of a fusion protein from a cell and cleavage of the secretory signal
 CC sequence from the fusion protein, are new. DNA encoding the Ytgss can be
 CC fused to either a reporter protein or a lipopolysaccharide-binding
 CC protein coding sequence. The isolated nucleic acid is useful for assaying
 CC for heterologous gene expression, detecting the presence of a compound
 CC that binds to an estrogen receptor in a sample or producing a desired
 CC protein from a host cell. It can also be used in a method for obtaining
 CC systemic circulation of a desired protein in a transgenic or chimeric
 CC host organism.
 CC
 SQ Sequence 51 AA;
 XX
 OY
 1 MRVLTALAVAVAGDSNMG 21
 |||||
 1 mrvltalavavagdsnmg 21
 Db
 RESULT 6
 AAB60944
 ID AAB60944 standard; Protein; 20 AA.
 XX
 AC AAB60944;
 XX
 DT 05-NOV-2001 (first entry)
 XX
 DE Horseshoe crab recombinant Factor C related peptide #9.
 XX
 KW Horseshoe crab; factor C; serine protease zymogen; endotoxin;
 KW coagulation cascade; bacterial infection; antibacterial.
 XX
 OS Unidentified.
 XX
 PN W0200127289-A2.
 XX
 PD 19-APR-2001.
 XX
 PF 13-OCT-2000; 2000WO-SG00162.
 XX
 PR 15-OCT-1999; 99US-0159569.
 PR 26-JUL-2000; 2000US-0626795.
 XX
 PA (UYST-) UNIV SINGAPORE NAT.
 XX
 PI Ding JL, Ho B, Tan NS;
 XX
 DR WPI; 2001-273778/28.
 PT Recombinant Factor C from *Carinoscorpius rotundicauda* is used to treat
 PT gram negative bacterial infections, detect the presence of gram
 PT negative bacteria and preserve samples from contamination by gram

PT negative bacteria -
 XX
 PS Disclosure; Fig 11B; 123pp; English.
 XX
 CC The present invention provides the protein and coding sequences of
 CC recombinant *Carinoscorpius rotundicauda* (the horseshoe crab) Factor C.
 CC Factor C is a serine protease zymogen activated by endotoxin (also known
 CC as lipopolysaccharide) to initiate the coagulation cascade, which is
 CC thought to be a defence mechanism of the organism against infection by
 CC gram negative bacteria. Factor C can be used to treat bacterial infection
 CC by binding to endotoxin and leading to bacteriostasis, having
 CC anti-endotoxic effects and microbicidal action against gram negative
 CC bacteria. It can also be used to detect the presence of endotoxin and
 CC develop systems to purify contaminated samples. The present sequence is
 CC a peptide used in the exemplification of the invention.
 CC
 SQ Sequence 20 AA;
 XX
 OY
 1 MRVLTALAVAVAGDSNMG 20
 |||||
 1 mrvltalavavagdsnmg 20
 Db
 RESULT 7
 AAR80624
 ID AAR80624 standard; Protein; 100 AA.
 XX
 AC AAR80624;
 XX
 DT 18-APR-1996 (first entry)
 XX
 DE Thiol protease inhibitor Virgiferin.
 XX
 KW Thiol protease inhibitor; virgiferin; Diabrotica virgifera; wCRM;
 KW western corn rootworm; digestive protease; papain-like; insecticide;
 KW Coleoptera; nematode infestation.
 XX
 OS Diabrotica virgifera.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..17
 FT Protein /label= signal_peptide
 FT 18..100
 FT /label= virgiferin
 FT "positions 64, 73 and 75 of the mature protein
 FT (i.e. positions 81, 90 and 92 using the
 FT numbering of the precursor, pre-virgiferin)
 FT are polymorphic sites; Pro or Leu residues
 FT can occur at position 64; Val or Leu
 FT residues can appear at position 73; and Pro
 FT or Arg can appear at position 75"
 XX
 PN W09524479-A1.
 XX
 PD 14-SEP-1995.
 XX
 PF 09-MAR-1995; 95WO-EP00881.
 XX
 PR 10-MAR-1994; 94US-0208571.
 XX
 PA (SANO) SANDOZ LTD.
 PA (SANO) SANDOZ PATENT GMBH.
 PA (SANO) SANDOZ-ERFINDUNGEN VERW GES MBH.
 XX
 PI Brunke KJ, Chan VJ, Deluca-Flaherty CR, Scarafra LEC;
 XX
 DR WPI; 1995-328271/42.
 DR N-PSDB; AAO98935.

XX Thiol protease inhibitor peptide, virgiferin, and its DNA - used to
 PT generate transformed plants with decreased susceptibility to damage
 PT by insect pests and nematodes

XX Claim 1; Page 25; 41pp; English.

XX The present sequence is that of the novel thiol protease inhibitor
 CC designated "virgiferin" that was isolated from Diabrotica virgifera
 CC (western corn rootworm). Virgiferin is a potent inhibitor of
 CC Diabrotica digestive proteases. Modified virgiferin peptides of
 CC mol. wt. 10-15 kDa (SDS-PAGE) with binding affinity to papain type
 CC proteases are also included. The cDNA coding for virgiferin can be
 CC used for generating transgenic plants able to produce the inhibitor;
 CC such plants will have decreased susceptibility to damage by insect
 CC pests, esp. Coleoptera, as well as to nematode infestation.

XX Sequence 100 AA;

Query Match 46.3%; Score 44; DB 16; Length 100;

Best Local Similarity 58.8%; Pred. No. 27;

Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 RVLVLAVALAVADPOS 18

Db 4 kvfvalavavavastqt 20

RESULT 8

AAAM19718 standard; Protein: 100 AA.

XX AAM19718;

XX 20-AUG-1997 (first entry)

XX Virgiferin.

XX Virgiferin; thiol protease inhibitor; western corn rootworm; insecticide;
 KW insect pest; nematode pest; coleopteran insect.

XX Diabrotica virgifera.

XX Key Location/Qualifiers

FT Peptide 1..17

FT Protein /note= "signal peptide"

FT Protein /note= "mature virgiferin"

XX US5629469-A.

XX 13-MAY-1997.

XX 10-MAR-1994; 94US-0208571.

XX 10-FEB-1995; 95US-0384367.

XX 10-MAR-1994; 94US-0208571.

XX (SANO) SANDOZ LTD.

XX Brunke RJ, Chan VJ, Deluca-Flaherty C, Scarafia LEC;

XX WPI: 1997-280334/25.

XX N-PSDB: AAT68798.

XX Thiol protease inhibitor virgiferin - useful as insecticide, and for

XX producing pest-resistant transgenic plants

XX Claim 14; Column 15-16; 17pp; English.

XX This sequence represents virgiferin. Virgiferin is a thiol protease

XX inhibitor isolated from the gut of Diabrotica virgifera (commonly known

CC as western corn rootworm). This protein is highly insecticidal, the
 CC mortality of D. virgifera larvae fed on a diet containing 4% of the
 CC protein was 100% on day six. This sequence, and modified versions of it,
 CC are useful as insecticides, especially for controlling Diabrotica spp.
 CC The DNA encoding this sequence is useful for the production of transgenic
 CC plants with reduced susceptibility to damage by pests that have thiol
 CC proteases as digestive enzymes, including insect and nematode pests,
 CC especially coleopteran insects.

XX Sequence 100 AA;

Query Match 46.3%; Score 44; DB 18; Length 100;

Best Local Similarity 58.8%; Pred. No. 27;

Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 RVLVLAVALAVADPOS 18

Db 4 kvfvalavavavastqt 20

RESULT 9

AAAM35283 standard; Protein: 381 AA.

XX AAM35283;

XX 17-FEB-1998 (first entry)

XX Human acid sphingomyelinase mutant fsp330.

XX Prenatal diagnosis; Type A; Type B; Niemann-Pick disease;
 KW identification; potential genetic transmitter; detection;
 KW recessive mutation; acid sphingomyelinase; Ashkenazi Jew;
 KW human; treatment; mutant frame shift Pro330.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Region 86..88

FT Region /note= "potential N-glycosylation site"

FT Region 175..177

FT Region /note= "potential N-glycosylation site"

XX US5686240-A.

XX 11-NOV-1997.

XX 27-MAY-1994; 94US-0250740.

XX 27-MAY-1994; 94US-0250740.

XX 03-MAY-1991; 91US-0695572.

XX (MOON) MOUNT SINAI SCHOOL MEDICINE.

XX Desnick RJ, Schuchman EH;

XX WPI: 1997-558133/51.

XX N-PSDB: AAT95068.

XX Diagnosing Type A or B Niemann-Pick disease - by detecting recessive

XX mutation in acid sphingomyelinase gene

XX Disclosure; Column -: 58pp; English.

XX Diagnosing Type A or B Niemann-Pick disease (NPD), or identifying a

XX person as having the potential to genetically transmit Type A or B

XX NPD, comprises detecting a recessive mutation in the acid

XX sphingomyelinase (ASM) gene, which results in an alteration of at

XX least 1 amino acid in the ASM amino acid sequence. The method is

XX especially useful for prenatal diagnosis in Ashkenazi Jewish

XX populations. The mutation is Arg949Glu, deltaArg608 Leu302Pro or

XX the fsp330 mutation described by the present sequence, where fsp330

CC and sequencing the amplified DNA or subjecting it to a
 CC hybridisation assay using mutation specific probes. The ASM type 1
 CC sequence, or the cDNA sequence encoding it can also be used in the
 CC treatment of NPD.
 CC N.B. Sequence not given in the specification, but constructed using
 CC the wild type ASM sequence given in columns 56-60.

XX
 SO Sequence 628 AA:

Query Match 46.3%; Score 44; DB 18; Length 628;
 Best Local Similarity 71.4%; Pred. No. 1.8e+02;
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 VLVLALVALAVGD 16
 DB 36 vlalalalalsd 49

RESULT 12

AA30642
 ID AAR30642 standard; Protein; 629 AA.

XX AAR30642;

DT 06-MAY-1993 (first entry)

DE ASM protein.

XX Acid sphingomyelinase; ASM; PASM-1FL; recombinant DNA; R496L; NPD;
 KM deltar608; delta1302; Neimann-Pick disease; Jewish community.

XX Homo sapiens.

OS
 XX
 FH Key Location/Qualifiers
 FT Peptide 363..419
 FT Peptide /note= "Type 1 sequence"

FT Peptide 83..92
 FT Peptide /note= "Tryptic peptide, T1"

FT Peptide 93..98
 FT Peptide /note= "Tryptic peptide, T2"

FT Peptide 239..249
 FT Peptide /note= "Tryptic peptide, T3"

FT Peptide 250..255
 FT Peptide /note= "Tryptic peptide, T4"

FT Peptide 295..304
 FT Peptide /note= "Tryptic peptide, T5"

FT Peptide 340..347
 FT Peptide /note= "Tryptic peptide, T6"

FT Peptide 349..359
 FT Peptide /note= "Tryptic peptide, T7"

FT Peptide 363..376
 FT Peptide /note= "Tryptic peptide, T8"

FT Peptide 377..381
 FT Peptide /note= "Tryptic peptide, T9"

FT Peptide 434..341
 FT Peptide /note= "Tryptic peptide, T10"

FT Peptide 541..555
 FT Peptide /note= "Tryptic peptide, T11"

FT Peptide 590..600
 FT Peptide /note= "Tryptic peptide, T12"

FT Peptide 87..88
 FT Peptide /note= "Potential N-glycosylation site"

FT Peptide 176..177
 FT Peptide /note= "Potential N-glycosylation site"

FT Peptide 336..337
 FT Peptide /note= "Potential N-glycosylation site"

FT Peptide 396..397
 FT Peptide /note= "Potential N-glycosylation site"

FT Peptide 504..505
 FT Peptide /note= "Potential N-glycosylation site"

FT Peptide 521..522
 FT Peptide /note= "Potential N-glycosylation site"

FT Misc-difference 322
 FT /note= "Amino acid varies from that found in
 FT Fibroblast cDNA's PASM-1 and PASM-2"

FT Misc-difference 506
 FT /note= "Amino acid varies from that found in
 FT Fibroblast cDNA's PASM-1 and PASM-2"

XX
 PN EP520843-A.

XX 30-DEC-1992.

XX 30-APR-1992; 92EP-0401241.

XX 03-MAY-1991; 91US-0695472.

XX (MOUN) MOUNT SINAI MEDICAL CENT.

XX Desnick RJ, Schuchman EH;

XX WPI; 1993-001632/01.

XX N-PSDB; AA033390.

XX Pure and recombinant acid sphingomyelinase and its nucleic acid
 XX for treatment and diagnosis of Niemann-Pick disease

XX Claim 14; Fig 3; 50pp: English.

XX This sequence represents functional acid sphingomyelinase (ASM) and
 XX was encoded by plasmid PASM-1FL. The nucleotide sequence encoding
 XX this protein may be used to generate recombinant DNA molecules that
 XX direct expression of the enzyme product. Certain mutations in the
 XX ASM gene ie. R496L, deltar608 and L302 have been found to correlate
 XX with Neimann-Pick disease (NPD). See also AA033391-423.

XX Sequence 629 AA:

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 46.3%; Score 44; DB 14; Length 629;
 Best Local Similarity 71.4%; Pred. No. 1.8e+02;
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 VLVLALVALAVGD 16
 DB 36 vlalalalalsd 49

RESULT 13

AA30643
 ID AAR30643 standard; Protein; 629 AA.

XX AAR30643;

DT 06-MAY-1993 (first entry)

DE R496L ASM.

XX Acid sphingomyelinase; ASM; type: 1; 2: PCR; primer: amplify; cryptic;
 KM polymerase chain reaction; splice site; mutation; R496L; deltar608;
 KW L302; Neimann-Pick disease; NPD; Jewish community.

XX Homo sapiens.

OS

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 46.3%; Score 44; DB 14; Length 629;
 Best Local Similarity 71.4%; Pred. No. 1.8e+02;
 Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 VLVLALVALAVGD 16
 DB 36 vlalalalalsd 49

RESULT 13

AA30643
 ID AAR30643 standard; Protein; 629 AA.

XX AAR30643;

DT 06-MAY-1993 (first entry)

DE R496L ASM.

XX Acid sphingomyelinase; ASM; type: 1; 2: PCR; primer: amplify; cryptic;
 KM polymerase chain reaction; splice site; mutation; R496L; deltar608;
 KW L302; Neimann-Pick disease; NPD; Jewish community.

XX Homo sapiens.

OS

XX

XX

XX

XX

XX

XX

XX

XX

```

XX (MOUN ) MOUNT SINAI MEDICAL CENT.
PA Desnick RJ, Schuchman EH;
PI WPI: 1993-001632/01.
XX N-PSDB; AAQ33393.
DR
XX
XX Pure and recombinant acid sphingomyelinase and its nucleic acid -
PT for treatment and diagnosis of Niemann-Pick disease
XX
XX Claims' 4 and 19; Fig 3; 50pp; English.
PS
XX This sequence represents the acid sphingomyelinase (ASM) mutation
CC R496L. The gene sequence encoding this protein was isolated by PCR
CC using primers constructed from common exonic sequences flanking the
CC type 1 and 2 specific sequences. Certain mutations in the ASM gene
CC ie. R496L, deltaR608 and L302 have been found to correlate with
CC Niemann-Pick disease (NPD). See also AAQ33390-423.
XX
XX Sequence 629 AA;
SQ
Query Match 46.3%; Score 44; DB 14; Length 629;
Best Local Similarity 71.4%; Pred. NO. 1.8e+02;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Oy 3 VLVIALAVLAAGD 16
Db 36 VIALALALALALSD 49
|||:||||:|
|:|:|:|:|:|:|

RESULT 14
AAR30645
ID AAR30645 standard; Protein: 629 AA.
XX
AC AAR30645;
XX
XX 06-MAY-1993 (first entry)
XX
XX L302P ASM.
XX
XX Acid sphingomyelinase; ASM; type: 1; 2; PCR; primer; amplify; cryptic;
XX polymerase chain reaction; splice site; mutation; R496L; deltaR608;
XX L302; Niemann-Pick disease; NPD; Jewish community.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX MISC-difference 302
XX /label- L302P
XX
XX EP520843-A.
XX
XX 30-DEC-1992.
XX
XX 30-APR-1992; 92EP-0401241.
XX
XX 03-MAY-1991; 91US-0695472.
XX
XX (MOUN ) MOUNT SINAI MEDICAL CENT.
XX
XX Desnick RJ, Schuchman EH;
XX
XX WPI: 1993-001632/01.
XX N-PSDB; AAQ33393.
XX
XX Pure and recombinant acid sphingomyelinase and its nucleic acid -
XX for treatment and diagnosis of Niemann-Pick disease
XX
XX Claims 6 and 21; Fig 3; 50pp; English.
XX
XX This sequence represents the acid sphingomyelinase (ASM) mutation

```

```

CC L302P. The gene sequence encoding this protein was isolated by PCR
CC using primers constructed from common exonic sequences flanking the
CC type 1 and 2 specific sequences. Certain mutations in the ASM gene
CC ie. R496L, deltaR608 and L302 have been found to correlate with
CC Niemann-Pick disease (NPD). See also AAQ33390-423.
XX
XX Sequence 629 AA;
SQ
Query Match 46.3%; Score 44; DB 14; Length 629;
Best Local Similarity 71.4%; Pred. NO. 1.8e+02;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Oy 3 VLVIALAVLAAGD 16
Db 36 VIALALALALALSD 49
|||:||||:|
|:|:|:|:|:|:|

RESULT 15
AAM35260
ID AAM35260 standard; Protein: 629 AA.
XX
AC AAM35260;
XX
XX 17-FEB-1998 (first entry)
XX
XX Human acid sphingomyelinase.
XX
XX Prenatal diagnosis; Type A; Type B; Niemann-Pick disease;
XX identification; potential genetic transmitter; detection;
XX recessive mutation; acid sphingomyelinase; Ashkenazi Jew;
XX human; treatment.
XX
XX Homo sapiens.
XX
XX Location/Qualifiers
XX Key
XX Region
XX /note- "potential N-glycosylation site"
XX 86..88
XX Region
XX /note- "potential N-glycosylation site"
XX 175..177
XX Region
XX /note- "potential N-glycosylation site"
XX 335..337
XX /note- "potential N-glycosylation site"
XX 395..397
XX /note- "potential N-glycosylation site"
XX 503..505
XX /note- "potential N-glycosylation site"
XX 520..522
XX /note- "potential N-glycosylation site"
XX Region
XX
XX US5686240-A.
XX
XX 11-NOV-1997.
XX
XX 27-MAY-1994; 94US-0250740.
XX
XX 27-MAY-1994; 94US-0250740.
XX
XX 03-MAY-1991; 91US-0695572.
XX
XX (MOUN ) MOUNT SINAI SCHOOL MEDICINE.
XX
XX Desnick RJ, Schuchman EH;
XX
XX WPI: 1997-558133/51.
XX N-PSDB; AAT95063.
XX
XX Diagnosing Type A or B Niemann-Pick disease - by detecting recessive
XX mutation in acid sphingomyelinase gene
XX
XX Claim 1; Columns 56-60; 58pp; English.
XX
XX Diagnosing Type A or B Niemann-Pick disease (NPD), or identifying a
XX person as having the potential to genetically transmit Type A or B
XX NPD, comprises detecting a recessive mutation in the acid

```

CC sphingomyelinase (ASM) gene, which results in an alteration of at
CC least 1 amino acid in the present ASM amino acid sequence. The
CC method is especially useful for prenatal diagnosis in Ashkenazi
CC Jewish populations. The mutation is Arg496Leu, deltaArg608,
CC Leu302Pro or TSP330, where TSP330 is a frame shift mutation
CC comprising a cytosine deletion in ASM codon 330. The mutations are
CC detected by selectively amplifying mutation containing portions of
CC the ASM gene by PCR using primers complementary and identical to a
CC portion of the cDNA sequence encoding the present sequence, and
CC sequencing the amplified DNA or subjecting it to a hybridisation
CC assay using mutation specific probes. The ASM type 1 sequence, or
CC the cDNA sequence encoding it can also be used in the treatment of
CC NPD.

XX
SQ Sequence 629 AA;

Query Match 46.3%; Score 44; DB 18; Length 629;
Best Local Similarity 71.4%; Pred. No. 1.8e+02;
Matches 10; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 3 VLVTALAVLAVGD 16
|||:|:|:
Db 36 v1a1a1a1a1sd 49

Search completed: August 6, 2002, 15:02:45
Job time: 210 sec

```
/organism="Danio rerio"  
/strain="AB"  
/db_xref="taxon:7955"
```



```
/clone="3817638"
/clone_lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/Note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTCG);
Site_2: DraIII (CACCATGTCG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGAGCCCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[GTGTGCTTACTGTCG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTCG, 3' site
CACCATGTCG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTTCGCTCTTAAAGCTGCG and 3' end
primer CGACCTGACGCTCAGCACA."
BASE COUNT      117 a      134 c      116 g      121 t      1 others
ORIGIN

Query Match      42.0%; Score 33.6; DB 10; Length 489;
Best Local Similarity 66.7%; Pred. No. 2.2;
Matches 48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 7 atccaccagcatgaggggtcgtgactgactgctgctgctgcagtgaggagacc 66
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 5 AACCCAGCAGCATGAGAGGCTGTGTGCTTGCCTGACTGTAGCCCTCGTGCGAGATCAAC 64

OY 67 agtccaactgg 78
    | | | | | | | |
DB 65 AGATGACACCTTG 76

RESULT 2
LOCUS      BP157180      502 bp      mRNA      linear      EST 23-FEB-2001
DEFINITION f115904.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
sequence.
3815983 5' similar to TR:093605 093605 VITELOGENIN ;, mRNA
sequence.
BP157180
ACCESSION  BP157180.1 GI:11052370
VERSION     EST.
KEYWORDS    zebrafish.
SOURCE      Danio rerio
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 502)
AUTHORS    Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
Washu zebrafish EST Project 1998
Unpublished (1998)
CONTACT    Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrfish@wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center Clone
distribution information can be found through the I.M.A.G.E.
Consortium/LBNL, send email to: info@image.llnl.gov
Seq primer: T3 RT from Amersham
High quality sequence stop: 489.
FEATURES
Source      1..502
Location/Qualifiers
```

```
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone="3815983"
/clone_lib="Sugano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/Note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTCG);
Site_2: DraIII (CACCATGTCG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGAGCCCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[GTGTGCTTACTGTCG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTCG, 3' site
CACCATGTCG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTTCGCTCTTAAAGCTGCG and 3' end
primer CGACCTGACGCTCAGCACA."
BASE COUNT      125 a      129 c      119 g      129 t
ORIGIN

Query Match      42.0%; Score 33.6; DB 10; Length 502;
Best Local Similarity 66.7%; Pred. No. 2.2;
Matches 48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 7 atccaccagcatgaggggtcgtgactgactgctgctgctgcagtgaggagacc 66
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3 AACCCAGCAGCATGAGAGGCTGTGTGCTTGCCTGACTGTAGCCCTCGTGCGAGATCAAC 62

OY 67 agtccaactgg 78
    | | | | | | | |
DB 63 AATTCACACCTTG 74

RESULT 3
LOCUS      AM133798      559 bp      mRNA      linear      EST 27-OCT-1999
DEFINITION f112a02.y1 Sugano Kawakami zebrafish DRA Danio rerio cDNA clone
sequence.
2600906 5' similar to TR:093605 093605 VITELOGENIN ;, mRNA
sequence.
AM133798
ACCESSION  AM133798.1 GI:6135405
VERSION     EST.
KEYWORDS    zebrafish.
SOURCE      Danio rerio
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 559)
AUTHORS    Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jos,S., Kucaba,T.,
Martin,J., Pape,D., Steptoe,M., Underwood,K., Theising,B., Ritter
,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
Washu zebrafish EST Project 1999
Unpublished (1999)
CONTACT    S.L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center
Seq primer: T3 RT from Amersham
High quality sequence stop: 527.
FEATURES
Source      1..559
Location/Qualifiers
/organism="Danio rerio"
```

```

/strain="AB"
/cb_xref="taxon:7955"
/clone="260906"
/clone_id="Sugano Kawakami zebrafish DNA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/notes="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer (ATGTGGCCCTTTTCTTTTCTTTT);
double-stranded cDNA was ligated to a DraIII adaptor
[GTGTGGCCCTGTG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTCTGCTCTTAAAGCTGCG and 3' end
primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      144 a      151 c      128 g      136 t
ORIGIN

```

```

Query Match      42.0%; Score 33.6; DB 9; Length 559;
Best Local Similarity 66.7%; Pred. No. 2.3;
Matches 48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

```

Qy 7 atccaccagcatgagggtgctgtactagctctgtgctgtgctgcagtgaggagacc 66
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4 AACCAACGACCATGAGACTGTGTGCTTCCCTTCAGCTGAGCCCTCGGGAGATCAAC 63
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 67 agtccactgg 78
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 64 AGATGACCTTG 75
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 4
LOCUS      B1876953      582 bp      mRNA      linear      EST 12-OCT-2001
DEFINITION f178f07.v1 Sugano Kawakami zebrafish DNA Danio rerio cDNA clone
            3820164 5' similar to SM:VITL_FUNHE Q90508 VITELLOGENIN I PRECURSOR
            ; mRNA sequence.
ACCESSION  B1876953
VERSION    B1876953.1 GI:16084224
KEYWORDS   EST.
SOURCE     zebrafish.
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE  1 (bases 1 to 582)
AUTHORS   Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
            Martin,J., Pape,D., Steptoe,M., Underwood,K., Theisling,B., Ritter
            ,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
            WashU Zebrafish EST Project 1999
            Unpublished (1999)
TITLE      Washington University School of Medicine
JOURNAL    4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
COMMENT    Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            Consortium/LLNL, send email to: info@genome.llnl.gov
            zebrafish identity (p-value greater than 1e-99) found to: TIGR:
            TC194 NUL
            High quality sequence stop: 520.
            Location/Qualifiers

```

FEATURES

```

source
1. .582
/organism="Danio rerio"
/strain="AB"
/cb_xref="taxon:7955"
/clone="3820164"
/clone_id="Sugano Kawakami zebrafish DNA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"
/lab_host="DH10B (phage resistant)"
/notes="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
with an oligo(dT) primer (ATGTGGCCCTTTTCTTTTCTTTT);
double-stranded cDNA was ligated to a DraIII adaptor
[GTGTGGCCCTGTG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
CACCATGTG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTCTGCTCTTAAAGCTGCG and 3' end
primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      146 a      137 c      133 g      166 t
ORIGIN

```

```

Query Match      42.0%; Score 33.6; DB 10; Length 582;
Best Local Similarity 66.7%; Pred. No. 2.3;
Matches 48; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

```

```

Qy 7 atccaccagcatgagggtgctgtactagctctgtgctgtgctgcagtgaggagacc 66
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 10 AACCAACGACCATGAGACTGTGTGCTTCCCTTCAGCTGAGCCCTCGGGAGATCAAC 69
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy 67 agtccactgg 78
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70 AGATGACCTTG 81
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

RESULT 5
LOCUS      B1876825      607 bp      mRNA      linear      EST 12-OCT-2001
DEFINITION f172a09.v1 Sugano Kawakami zebrafish DNA Danio rerio cDNA clone
            3819569 5' similar to TR:093605 093605 VITELLOGENIN I, mRNA
            sequence.
ACCESSION  B1876825
VERSION    B1876825.1 GI:16084096
KEYWORDS   EST.
SOURCE     zebrafish.
ORGANISM   Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
            ; Cyprinidae; Danio.
REFERENCE  1 (bases 1 to 607)
AUTHORS   Sugano,S., Kawakami,K., Johnson,S., Li,F., Marra,M., Eddy,S.,
            Hillier,L., Clifton,S., Allen,M., Gibbons,M., Jost,S., Kucaba,T.,
            Martin,J., Pape,D., Steptoe,M., Underwood,K., Theisling,B., Ritter
            ,E., Bowers,Y., Wylie,T., Waterston,R. and Wilson,R.
            WashU Zebrafish EST Project 1999
            Unpublished (1999)
TITLE      Washington University School of Medicine
JOURNAL    4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
COMMENT    Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
            Sequencing by: Washington University Genome Sequencing Center Clone
            Consortium/LLNL, send email to: info@genome.llnl.gov
            High quality sequence stop: 518.
            Location/Qualifiers

```

FEATURES

Fax: 314 286 1810
Email: zbrafish@watsen.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA Sequencing by Washington University Genome Sequencing Center Clome distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: T3 ET from Amersham
High quality sequence stop: 485.

FEATURES
source

```

1. 718
/organism="Danio rerio"
/strain="AB"
/db_xref="taxon:7955"
/clone="3815731"
/clone_lib="Susano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, including
unfertilized eggs)"
/dev_stage="adult"

```

/ab, host="DH10B (phage resistant)"
/note="vector: pME185-Fl3; site.1: DraIII (CACTGTG);
site.2: DraIII (CAACATGTG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGAGCGCTTTTCTTTTCTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TTCGGCTACTGG], digested and cloned into distinct DraIII
sites of the pME185-Fl3 vector (5' site CACTGTG, 3' site
CAACATGTG). XhoI should be used to isolate the cDNA
insert. Site selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer CTTTCGCTTAAAGCTGG and 3' end
primer GAACCTGCAGCTCGACCA.

BASE COUNT	194 a	183 c	169 g	171 t	1 others
ORIGIN					

Query Match	42.0%;	Score 33.6;	DB 10;	Length 718;
Best Local Similarity	66.7%;	Pred. No. 2.5;		
Matches 48;	Conservative	0;	Mismatches 24;	Indels 0;
			Gaps	0;

Oy 7 atccaccagccatgagggtgctgtactagtcttctgtgtcctgcgaagtgggacc 66
| | | | | | | | | | | | | | | | | |
Db 3 AACCAACGACATGAGAGCTGTTCGCTTCCCTGACTGTAGCCCTCGTGCGAGTCAAC 62

QY	67	agtc	caact	tg	78
Db	63	AGATGA	ACCTTC	74	

RESULT 10

BI875920

LOCUS	722 bp	mRNA	linear	EST 12-OCT-93
DEFINITION	fl64g01.y1	Sugano	Kawakami zebrafish DRB Danto retio cDNA clone	
	3818688 5'	similar to TR:093605	093605 VITELLOGENIN ; mRNA	

ACCESSION	BI875920
VERSION	BI875920.1
GI	GI:16083191

SOURCE	zebrafish.
ORGANISM	Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE AUTHORS

Martin, J., Pape, D., Steptoe, M., Underwood, K., Theising, B., Ritter, E., Bowers, Y., Wylie, T., Waterston, R. and Wilson, R.

TITLE	Washu zebrafish EST Project 1999
JOURNAL	Unpublished (1999)

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami
Sequencing by Washington University Genome Sequencing Center Clones
Distribution information can be found through the I.M.A.G.E.
Consortium/LNU, send email to: info@imgc.lnu.gov
High quality sequence stop: 519.

FEATURES

Source

```
1..722
/organism="Danio rerio"
/strain="Ab"
/db_xref="taxon:795"
/clone="3818688"
/clone_lib="Susano Kawakami zebrafish DRA"
/sex="mixed (one male and one female, unfertilized eggs)"
/dev_stage="adult"
```

/lab_host="DH10B (phage resistant)"
/note="Vector: pME18S-FLJ; Site.1: DraIII (CAGCTGTGG);
Site.2: DraIII (CACCATCGT); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTCGCCTTTTTTTTTTTTTTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TGTGGCCCTACTGG], digested and cloned into distinct DraIII
sites of the pME18S-FLJ vector (5' site CAGCTGTGG, 3' site
CACCATCGT). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science) and
kindly donated by Dr. Koichi Kawakami. Custom primers for
sequencing: 5' end primer GTTGTCCTTAAGAAGTCGG and 3' end
primer CGAATGCAGCTGCACCA."

BASE COUNT	191 a	185 c	172 g	171 t	3 others
ORIGIN					

Query Match	42.08;	Score 33.6;	DB 10;	Length 722;
Best Local Similarity	66.78;	Pred. No. 2.5;		
Matches	48;	Conservative	0;	Mismatches 24;
			Indels 0;	Gaps 0;

Oy 7 atccaccagcattgaaggctgtctgtactagctcttgcgtgacctcgcagttggagacc 66
| | | | | | | | | | | | | | | | | |
Db 1 AACCAACGACCATGAGAGCTGTGTGGCTTGCCCTGACTGTAGCCCTCGTGGGAGTCAC 60

QY	67	agtc	caact	tg	78
Db	61	AGATGA	ACCTTG	72	

RESULT 11

BI876481

LOCUS	BI876481	727 bp	mRNA	linear	EST 12-
DEFINITION	fl172h10.y1	Sugano	Kawakami	zebrafish	DRA Danio rerio cdna c
	3819522 5'	similar to	TR:093605	093605	VITELOGENIN ;, mRNA

ACCESSION	BI876481
VERSION	BI876481.1
	GI:16083752

SOURCE	zebrafish.
ORGANISM	Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
Cyprinidae; Danio.

REFERENCE
AUTHORS

Wallerstein, J., Martin, J., Pape, D., Steptoe, M., Underwood, K., Theising, B., Ritten, E., Bowers, Y., Wylie, T., Waterston, R. and Wilson, R.

TITLE	JOURNAL
washu zebrafish EST Project 1999	
Unpublished (1999)	

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

